FS1

English Translation of Japanese Publication

08/951,733

A-433B

WORLD ORGANIZATION FOR INTELLECTUAL PROPERTY

International patent published on the basis of Patent Cooperation Treaty

(11) International Publication No. **WO 98/07838**

(43) International Publication Date: February 26, 1998

(51) International Patent Classification[6]: C12N 9/12, 15/54, C12Q 1/48

(21) International Patent Application No.: PCT/JP97/02904

(22) International Application Date: August 21, 1997

(30) Priority Data

Priority No.: Hei 8[1996]/219761

Priority Date: August 21, 1996

Priority Country: JP

Priority No.: Hei 9[1997]/18878

Priority Date: January 31, 1997

Priority Country: JP

Priority No.: Hei 9[1997]/31807

Priority Date: February 17, 1997

Priority Country: JP

(71) Patent Applicant (for all contracted nations except USA):

Mitsubishi Chemical Corp.

2-5-2 Marunouchi, Chiyoda-ku, Tokyo (JP)

(72) Inventor(s) and

(75) Inventor(s)/Applicant(s) (USA only):

Fuyuki Ishikawa

3-16-105 Shinoharadai-machi, Minatokita-ku, Yokohama-shi, Kanagawa-ken (JP)

Hideo Nakamura (JP/JP)

Kazuhiro Takahashi (JP/JP)

Yasuhiro Fujino (JP/JP)

Naozumi Harada (JP/JP)

Mitsubishi Chemical Corp.

Yokohama General Research Center, Kanagawa (JP)

1000 Kamoshida-cho, Aogusa-ku, Yokohama-shi Kanagawa-ken

(74) Representative:

Masazumi Imamura et al., Patent Attorney

Touwa Yaesu 1-chome Bldg. 7th floor

8-12 Yaesu 1-chome, Chuo-ku, Tokyo

(81) Contracting Nations:

CA, JP, US, European Patents (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Appended Kokai Documents: International Search Document

(54) Title:

HIGHER ANIMAL TELOMERASE PROTEIN AND GENE ENCODING THE SAME

(57) Abstract:

A telomerase protein and gene encoding the same are offered that originate in higher animals including humans. The telomerase protein and gene encoding the same are expected to be useful in understanding biological control mechanisms such as cellular growth and senescence, and are expected to be particularly useful in the development of cancer therapies. In addition, a screening method is also offered for screening substances that act on the expression of the enzymatic activity of higher-animal telomerase protein, where this screening method includes a process wherein SDS polyacrylamide electrophoresis is used in the measurement of the molecular weight of telomerase protein present in cells or tissues in contact with a test substance.

HIGHER ANIMAL TELOMERASE PROTEIN AND GENE ENCODING THE SAME

Technical Field

The present invention concerns a gene that codes for telomerase in higher animal cells, and the gene product thereof.

Technological Background

The ends of linear DNA in the chromosomes of eukaryotic cells such as animal cells are referred to as telomers, and these ends have a complicated higher-order structure consisting of DNA sequences and proteins that bind these sequences. Telomeric DNA is composed of a characteristic repeating sequence rich in thymine (T) and guanine (G) (or adenine (A) and cytosine (C) for the opposite strand), and for example, the telomeric DNA of vertebrate cell chromosomes is composed of a repeat of 6 bases TTAGGG (or CCCTAA for the opposite strand). The average length of the telomeric repeats in human cells, determined by analyzing this sequence by southern blotting, has been shown to be 7-10 kb.

Telomer structure is thought to have an important function in the stabilization of chromosomes. For example, it has been shown by morphological studies in yeast that the location of telomers at the periphery of the cell nucleus allows them to function as anchors for fixing the chromosomes at specific locations in the nucleus, and it has been suggested that telomers might regulate physical crossing-over between chromosomes within the cell nucleus. In addition, as will be described below, it has been suggested that telomers have the function of preventing deactivation of chromosome function due to shortening occurring with each replication of linear double-stranded DNA in eukaryotic cells.

In the process where simultaneous replication of both strands of linear duplex DNA occurs, one of the DNA strands (the leading strand) is continuously replicated by a polymerase in the 5' to 3' direction using the 3' terminal as a primer, whereas the other DNA strand (lagging strand) is discontinuously replicated using RNA primers.

Consequently, the RNA primer at the 5' terminal of the newly formed strand (lagging strand) cannot be converted into DNA, so that the 5' end of each successive daughter cell gradually shortens with repetition of cellular division. Eventually, the chromosome will become unstable, leading to cell death. However, it has been shown that shortening of chromosomal DNA and loss of chromosome function does not occur with repeated DNA replication in germ line cells (Allsopp R.C. et al., Proc. Natl. Acad. Sci. USA 89, 10114, 1992), and it has been suggested that it is possible that the telomers or regions adjacent thereto assume a hair-pin structure, thereby functioning as a buffer zone against shortening.

The function of telomers in preventing chromosomal shortening is strongly suggested based on the relationship between the variation in the average length of telomer repeat sequences and cellular senescence and death. When fibroblast cells from multicellular organisms are cultured *in vitro* by passaging, their proliferative capacity decreases with each successive passage until finally they become senescent cells with no proliferative capacity. However, there are cases in which immortal cells are obtained that are endowed with an unlimited capacity for proliferation when certain cancer genes are introduced into the cells. Although these cells are used as models for studying the phenomenon of senescence and carcinogenesis on the cellular level (*in vitro*), it has been found, based on research at the molecular level, that the average length of telomeric repeat sequences decreases with increasing numbers of cell divisions for normal cells, and that this average length is related to the potential number of passages. Moreover, it has been found that there is no change in average length during passaging with immortal cells, although the average length of their telomeric repeat sequences may be short.

RNA-dependent DNA polymerase (telomerase) that lengthens the telomeric repeat sequences has received attention as one mechanism for regulating the average length of telomeric repeat sequences. This enzyme was discovered as an enzyme that adds the same 6-base repeat sequence to the 3' end of a synthetic oligonucleotide (TTGGGG) from the tetrahymena telomeric repeat sequence obtained using macronucleus extract from the protozoan tetrahymena. This enzyme is a type of reverse transcriptase that contains, as a subunit that is required for its activity, template RNA that

is complementary to the 5'-TTAGGG-3' of the telomeric DNA sequence. The enzyme lengthens a single strand of the telomeric DNA based on this template RNA. Telomerase has been purified from tetrahymena telomerase, and its cDNA has been cloned (Collings K. et al., Cell, 81, 677, 1995). This telomerase is composed of a 95 kDa subunit that binds the DNA terminal serving as a primer, and an 80 kDa subunit that binds template RNA. It was shown that this telomerase has a primary structure that is somewhat similar to the RNA polymerases of RNA viruses.

The biological significance of telomerase has been demonstrated in lower eukaryotes such as tetrahymena and yeast. Specifically, with cells that are transformed with genes having point mutations in the template region of the telomeric repeat sequence of the tetrahymena telomerase RNA gene, proliferation becomes impossible in conjunction with biosynthesis of the mutated telomer repeat sequence including the introduced point mutation. In addition, when TLC1, the telomerase RNA gene of baker's yeast, is disrupted, the average length of the telomeric repeat sequences in the yeast shortens along with repeated passaging, and loss of proliferative capacity eventually occurs. Telomerase is thus understood to be an enzyme that is required for cellular proliferation in unicellular eukaryotes.

In the senescence process occurring *in vitro* in human cells, telomerase activity has not been observed in the initial period of passaging after introduction of cancer genes, but this activity has been detected in cell lines that have a capacity for infinite reproduction. In addition, although telomerase activity has been detected in almost all actual human cancer cells, it is reported that telomerase activity has not been detected in most normal cells. Based on this information, the supposition can be put forth that it may be possible to endow cells with infinite proliferative capacities by preventing the shortening of telomeric DNA by means of the expression of telomerase activity. Consequently, telomerase inhibitors would be useful as highly selective anticancer agents, and the possibility of an early diagnosis of cancer by means of a telomerase activity assay has been projected.

It has been reported that the degree of telomerase RNA subunit expression is not necessarily related to telomerase activity (Avilon et al. Cancer Res., 56, 645, 1996).

However, telomerase itself has not yet been isolated and purified from higher animals including humans, and so the physical reality is unclear at present. Moreover, a complicated assay involving the use of PCR is required in order to detect actual telomerase activity, and at present, there are almost no enzymatic studies of telomerase in existence. Furthermore, determining whether there is a positive relationship between cancer malignancy and telomerase is difficult because the expression of telomerase cannot be detected at the level of the individual cell using pathologic sections.

Consequently, there is a strong desire for the isolation and identification of telomerase protein, which will lead to the clarification of the physical characteristics of higher animal telomerase, research into telomerase inhibitors based on enzymatic knowledge, and clarification of the relationship between telomerase and cancer malignancy.

Development of the invention

The inventors of the present invention et al. carried out painstaking investigations towards the isolation and identification of a higher animal telomerase protein. The inventors succeeded in cloning the gene encoding higher animal telomerase protein and in expressing the higher animal telomerase protein that is the gene product of this gene. In addition, the inventors succeeded in producing an antibody that specifically recognizes this gene product, and used this antibody to demonstrate a strong relationship between telomerase activity and this gene product. The present invention was perfected based on this knowledge. Recently, the entire amino acid sequence of human telomerase protein was published (Science, 275, pp.973-977, February 14, 1997), but this sequence differs in many regions from the base sequence and amino acid sequence of the c-DNA found by the present inventors et al.

The present invention offers a polypeptide that is specified by the amino acid sequence described in sequence no. 1 in the sequence table, where said polypeptide is characterized by being a telomerase protein derived from rat. The present invention offers a polypeptide that is characterized by functioning essentially as the telomerase protein for higher animals, including humans, where two or more substitutions,

insertions, and or deletions are present in the amino acid sequence of sequence no. 1 in the sequence table. In a preferred mode of the present invention, the aforementioned polypeptide is offered that can function as the telomerase protein in the human body.

In addition, in another mode of the present invention, a polypeptide is offered that is specified by the amino acid sequence represented by sequence no. 2 in the sequence table, but this polypeptide is characterized by being a partial polypeptide of the telomerase protein derived from humans. In addition, the present invention offers a polypeptide that is characterized by functioning essentially as a partial polypeptide of the telomerase protein of higher animals, including humans, which has one or more amino acid sequence substitutions, insertions and/or deletions in the amino acid sequence represented by sequence no. 2 in the sequence table.

In yet another mode of the present invention, a polypeptide is offered that is specified by the amino acid sequence represented by sequence no. 13 in the sequence table, where said polypeptide is characterized by being a telomerase protein derived from humans. In addition, the present invention offers a polypeptide that is characterized by functioning essentially as the telomerase protein of higher animals including humans, with one or more amino acid sequence substitutions, insertions and/or deletions in the amino acid sequence represented by sequence no. 13 in the sequence table. In a preferred mode thereof, the present invention offers the aforementioned polypeptide that can function as telomerase protein in the human body.

Another mode of the present invention offers the nucleotide sequences that encode the aforementioned various polypeptides. DNA sequences and RNA sequences can be offered as nucleotide sequences, and for example, in preferred modes, DNA is offered that is specified by the sequence spanning nucleic acid nos. 199-8085 (including the termination codon) in the DNA sequence represented by sequence no. 1 in the sequence table, DNA is offered that is specified by the sequence spanning nucleic acid nos. 1-487 in the DNA sequence represented by sequence no. 2 in the sequence table, and DNA is offered that is specified by the sequence spanning nucleic acid nos. 156-8030 (not including the initiation codon) in the DNA sequence represented by sequence no. 13 in the sequence table. In addition to the above, recombinant vectors containing the

aforementioned DNA sequences, transformants containing said recombinant vector, and a method for manufacturing the aforementioned polypeptides that includes the process for isolating and recovering the polypeptides that are the gene products of the aforementioned DNA sequences from cultures of said transformants, are offered.

Another mode of the present invention offers nucleic acid probes that include nucleotide sequences that can bind complementary to part or all of the aforementioned nucleotide sequences and antibodies that can specifically recognize the aforementioned various polypeptides. These antibodies or nucleic acid probes are useful as reagents for detecting cancer cells, and a medical composition for use in cancer diagnosis that includes the aforementioned antibodies or nucleic acid probes is also offered as a mode.

In addition to these inventions, a further mode of the present invention offers the aforementioned polypeptide, characterized by having an inactive form with a molecular weight of about 240 kDa and an active form with a molecular weight of about 230 kDa, as determined by SDS (sodium dodecylsulfate) polyacrylamide electrophoresis (PAGE), and an active polypeptide, characterized by having a molecular weight of about 230 kDa as determined by SDS polyacrylamide electrophoresis. A process is also offered that pertains to a screening method for substances that act on the expression of the enzymatic activity of higher animal telomerase protein, where the molecular weights of the polypeptides that are the subunits of higher animal telomerase protein contained in cells or tissues that are in contact with a test substance are measured.

Preferred modes of the invention pertaining to the method described above offer the aforementioned method wherein the process that involves contact with the test substance is performed by culturing in the presence of the test substance, or by administration of the test substance to animals; the aforementioned method wherein measurement of the molecular weight is carried out by SDS polyacrylamide electrophoresis; the aforementioned method that includes a process wherein the ratio of the approximately 240 kDa inactive polypeptide and the approximately 230 kDa active polypeptide is measured; the aforementioned method that includes a process wherein it is determined that said test substance inhibits the expression of enzymatic activity of higher animal telomerase protein if there is an essential increase in the ratio of 240 kDa

polypeptide in the presence of the test substance relative to the ratio of said polypeptide in the absence of the test substance; and the aforementioned method that includes a process wherein it is determined that said test substance inhibits the expression of enzymatic activity of higher animal telomerase protein when there is an essential increase in the ratio of 230 kDa polypeptide in the presence of the test substance relative to the ratio of said polypeptide in the absence of the test substance.

Brief Description of the Figures

Figure 1 is a diagram showing a restriction enzyme cleavage site map of the cDNA clone of the rat telomerase protein gene.

Figure 2 is a diagram showing the results of comparing the homology of the DNA sequence obtained from a cDNA fragment of the human telomerase protein gene expanded by PCR, and the predicted amino acid sequence thereof, with the respective sequences for rat and tetrahymena p80. In the figure, R denotes the rat gene, H denotes the human gene, and p80 denotes the tetrahymena p80 gene.

Figure 3 is a diagram showing the results of immunoprecipitation of telomerase activity derived from extracts of rat cancer cells (AH66F) or human cancer cells (PA-1) using beads coated with antibodies specific for a fragment of recombinant rat telomerase protein. The results of assay using a method that combines PCR and ELISA are shown, where the vertical axis represents the telomerase activity. "Beads alone" denotes a negative control where antibodies were not coated onto the beads, "PI-1" denotes a negative control where IgG derived from pre-immunized serum was coated, and "1-41d" and "R1-116d" show the results of samples produced by coating specific IgG derived from hyperimmunized serum.

Figure 4 is a diagram showing a restriction enzyme cleavage site map of the cDNA clone of the human telomerase protein gene.

Optimal Mode for Implementing the Invention

The first mode of the polypeptide of the present invention corresponds to a polypeptide that constitutes telomerase protein derived from mouse, and is specified by

the amino acid sequence expressed by sequence no. 1 in the sequence table. The aforementioned polypeptide that is offered by the present invention is not limited to the polypeptide that is specified by sequence no. 1. Polypeptides that have one or more amino acid residue substitutions, insertions and/or deletions in the amino acid sequence represented by sequence no. 1 in the sequence table, and that can function essentially as telomerase protein in higher animals including humans, are also within the scope of the invention. Moreover, higher animal telomerase proteins that include this polypeptide as a subunit are also within the scope of the present invention.

The second mode of the polypeptide of the present invention corresponds to a partial polypeptide of the polypeptide that constitutes the telomerase protein derived from humans, and is specified by the amino acid sequence represented by sequence no. 2 in the sequence table. The aforementioned polypeptide that is offered by the present invention is not limited to the polypeptide that is specified by sequence no. 2. Polypeptides that have one or more amino acid residue substitutions, insertions and/or deletions in the amino acid sequence expressed by sequence no. 2 in the sequence table, and that can function essentially as partial polypeptides of higher animal telomerase protein, and preferably human telomerase protein, are also within the scope of the invention.

The third mode of the polypeptide of the present invention corresponds to a polypeptide that constitutes human telomerase protein, and is specified by the amino acid sequence represented by sequence no. 13 in the sequence table. The aforementioned polypeptide that is offered by the present invention is not limited to the polypeptide that is specified by sequence no. 13. Polypeptides that have one or more amino acid residue substitutions, insertions and/or deletions in the amino acid sequence represented by sequence no. 13 in the sequence table, and that can function essentially as telomerase protein in higher animals including humans, are also within the scope of the invention. Moreover, higher animal telomerase proteins that include this polypeptide as a subunit are also within the scope of the present invention.

The polypeptide of the present invention also includes polypeptides that contain the aforementioned various polypeptides as partial sequences. For example, a polypeptide comprising any of the aforementioned polypeptides linked to an appropriate

amino acid sequence that has the property of upregulating expression efficiency, a polypeptide comprising a signal sequence linked to any of the aforementioned polypeptides, or a so-called fusion protein with a so-called tag sequence comprising the aforementioned polypeptide linked to another protein in such a manner that the reading frame is unaltered, thereby ensuring expression of the aforementioned polypeptide, are all within the scope of the invention.

The nucleotide sequences coding for any of the aforementioned polypeptides are all included in the nucleotide sequences of the present invention. The nucleotide sequences (preferably DNA sequences) that encode the polypeptides included in the first mode, second mode and third mode described above are instances of the genes that encode for the telomerase protein of the present invention (in this specification, the term "telomerase protein gene" means the nucleotide sequences that code for all or part of the length of the polypeptides that constitute the telomerase protein).

In this specification, the term "higher animal" is a general term that includes mammals such as humans. The polypeptides that constitute the telomerase proteins derived from higher animals, and preferably mammals, are expected to have high homology. Consequently, it is obvious that, based on the genetic information and the cloning methods for the telomerase protein gene derived from mice as specified in this specification, it would be possible for a person skilled in the art readily to obtain the genes coding for the polypeptides that constitute the telomerase proteins derived from higher animals, and to thereby obtain the gene products.

The telomerase protein gene of the present invention is obtained, for example, by the method described below. A plasmid cDNA library, phage cDNA library or phage genomic library produced by common well-known methods using RNA prepared from immortalized higher animal cell lines such as human, monkey, horse, sheep, pig, cat, rabbit, rat or mouse cell lines can be employed as the DNA library containing the telomerase protein gene of the present invention.

For example, when a phage cDNA library is used, first, cancer or other tissue, or an immortalized higher animal cell line is pulverized in liquid nitrogen, and is homogenized in guanidine isothiocyanate aqueous solution. Following the method of

Chirgwin et al. (Biochemistry 18, 5294-5299 (1979)), the total RNA fraction is then separated as precipitate by cesium chloride equilibrium density gradient centrifugation. An extraction reagent such as the commercially-available RNAzol (Tel Test) can also be used in separating the RNA. After separating the RNA, the total RNA is purified by phenol extraction and ethanol precipitation, and is then further purified using oligo-(dT) cellulose column chromatography to prepare an mRNA (poly-$(A)^+$ mRNA) group that includes the mRNA of the target telomerase protein.

Next, for example, an oligo-(dT) sequence consisting of 12-18 deoxythymidines or primer DNA composed of synthetic DNA containing an oligo-(dT) sequence is hybridized to the mRNA group prepared as described above, and single stranded cDNA is synthesized using reverse transcriptase as described in Nature 329, 836-838 (1987). Sequences similar to these are used in commercially available cDNA synthesis kits, and so this type of sequence can also be used. A PCR reaction is then carried out using synthetic DNA (ordinarily, material included in the kit) for PCR reactions with a commercially available primer. In addition, when primer DNA is used as described in the aforementioned publication (Nature 329, 836-838 (1987)), a sequence homologous to this sequence is designed, and can be prepared beforehand for use as primers in the PCR reaction. E. coli DNA polymerase I, E. coli DNA ligase and RNase H are then used in the synthesis of double-stranded cDNA by a common method. The cDNA ends are then blunted using T4 DNA polymerase, and short fragments of DNA for producing a form that can be cut with restriction enzymes, so-called EcoRI adapters, are attached to both ends of the cDNA using T4 DNA ligase.

Similar results can be obtained, at this time, by methylating the cDNA restriction enzyme cleavage sites using a DNA methylase such as EcoRI methylase (with EcoRI methylase, for example, methylation occurs at the EcoRI restriction site), thereby protecting the cDNA from cutting by the restriction enzyme EcoRI. Next, so-called EcoRI linkers, etc., are attached to the ends of the cDNA using T4 DNA ligase, so that only the linker DNA region is cut by the restriction enzyme EcoRI. When the cleavage site of another restriction enzyme such as Bam HI is selected as the vector cloning site, similar results can be obtained by carrying out the aforementioned series of terminal

treatment processes involving, for example, linkage to a Bam HI adaptor or treatment using a combination of Bam HI methylase, Bam HI linkers and Bam HI.

The cDNA that has been subjected to terminal treatment as described above is then inserted at the EcoRI cleavage site of a commercially available λ phage vector, such as λ ZAP (Promega Biotech) or other λ phage vector, or a plasmid vector such as pGEM2 (Promega Biotech) according to a common method, thereby producing a recombinant plasmid DNA group or a recombinant λ phage group. Alternately, when a PCR reaction is used to produce fragments, an (A) is specifically attached to the terminals of the DNA fragments that have been expanded by PCR, and thus manufactured can be carried out using a vector that has a corresponding T, such as pCRII (Invitrogen) or pT7 (Novagen).

The recombinant λ phage DNA group obtained in this manner is then used, and so-called *in vitro* packaging is carried out with, for example, a commercially available *in vitro* packaging kit such as Gigapack Gold (Promega Biotech). λ phage particles can thus be produced that contain the recombinant λ phage DNA. Packaging is generally carried out under the conditions prescribed in the pamphlets accompanying commercially available kits. The resulting λ phage particles are then transformed into a host such as E. coli according to a common method such as the method described by Maniatis et al. ("Molecular Cloning", Cold Spring Harbor Laboratories, 1982). The resulting transformants are then grown in order to produce a phage cDNA library. In addition, when a recombinant plasmid DNA group is used, this group can be used in the transformation of a host such as E. coli, and a plasmid cDNA library can be obtained by growing the resulting transformants.

Next, these phage or E. coli transformants can be grown, and transferred onto nitrocellulose film or nylon film such as Gene Screen Plus (DuPont), whereupon the protein can be removed in the presence of alkali. The λ phage DNA or plasmid DNA that has been prepared in his manner is then selected by a plaque hybridization method wherein a ^{32}P labeled probe produced from a partial fragment of the higher animal telomerase protein gene expanded by the method described below is hybridized to the plasmid DNA or λ DNA. All or part of the target cDNA clone coding for the higher animal telomerase protein gene can thereby be obtained.

The probe that is used in the selection of the cDNA clone that codes for the higher animal telomerase protein gene that is used for targeting the plasmid cDNA library or phage cDNA library can then be prepared according to a common method using, for example, a commercially available kit. For example, a DNA sequence derived from the gene that codes for a known telomerase protein (Collins et al., Cell, 81, 677-686, 1995), or a DNA sequence from the gene of another animal that can code for an amino acid sequence that is homologous to this amino acid sequence, can be determined using TBLASTN or some other program from a gene bank such as the National Center for Biotechnology Information (NCBI). An amino acid sequence that has a certain level of homology can then be used as a basis for synthesizing oligonucleotide probes in reference to the DNA sequence coding for the amino acid sequence. These sequences can then be used as probes. In addition, PCR primers can be constructed based on the DNA sequence of a similar gene, and so-called degenerative [sic] PCR can be used in order to produce longer DNA sequences, which can then be used as probes. In this case, the template that is used in the PCR method can be a phage cDNA library or a plasmid cDNA library derived from cells that contain the DNA of the target probe, or cDNA synthesized from RNA extract according to a common method.

In addition, part of the higher animal telomerase protein gene can be obtained by a PCR method wherein PCR primers are designed so that probe DNA is produced without screening the gene library using hybridization as described above. In this case, the aforementioned phage cDNA library, plasmid cDNA library or other cDNA synthesized from RNA extracted from immortalized cells can be used directly as the template in the PCR method. After the PCR reaction, the reaction liquid is analyzed by agarose or polyacrylamide gel electrophoresis. Fragments of the predicted size thus can be obtained from among the DNA fragments that have been expanded by using two types of primers, and these fragments can be purified. For example, the fragments can be introduced into a commercial vector that can directly incorporate PCR fragments such as pCR-II, and a host such as E. coli can be transformed with the resulting recombinant vector, which can then be used in base sequence analysis. In addition, new PCR primers can be designed based on the partial sequence of the resulting higher animal telomerase protein gene,

these primers can then be synthesized, and the gene coding for the entire length of the higher animal telomerase protein can then be obtained by repeated expansion of the DNA between the aforementioned newly synthesized primer and PCR primers designed based on the sequence of the higher animal telomerase protein, primers having sequences that are homologous to the primers used in cDNA synthesis, PCR primers that correspond to the anchor sequences added to both ends of the cDNA, or primers for the vector in which the cDNA has been incorporated.

After completion of the PCR reaction, the DNA fragments can be analyzed, recovered and purified by a common method used in conjunction with agarose or polyacrylamide gel electrophoresis. The resulting purified DNA fragments are then inserted into a vector that can directly incorporate PCR fragments such as pCR-II, and E. coli can be transformed with the resulting recombinant vector. DNA can then be prepared following a common method, and can be sequenced by the dideoxy method of Sanger et al. (Proc. Natl. Acad. Sci. USA, 74, 5463, 1977) in order to determine the base sequence of the target DNA fragment.. Sequence determination can be carried out using an automatic sequencer such as ABI373A (Applied Bio Systems).

When a clone is obtained from a phage library or plasmid library, there are cases where it is difficult to analyze the entire region of the cDNA that has been inserted into the vector, because there are generally limitations on the sequence length for which the base sequence can be determined when an automatic sequencer is used. In such a case, the fragment is cut with the appropriate restriction enzyme, and the fragments are separated and recovered by gel electrophoresis. The recovered fragments can then be analyzed by re-insertion into an appropriate vector. In addition to this type of procedure (sub-cloning), an appropriate sequence can be selected from the base sequence that has been identified by an automatic sequencer, new primers can be designed, and sequencing can be continued from this point so that analysis is carried out discontinuously. By then linking the sequences of the DNA fragments that have been sequenced in this manner so that they overlap each other, a determination can be made as to the entire length of the nucleotide sequence coding for the polypeptides that constitute higher animal telomerase protein, as represented by sequence nos. 1-13 in the sequence table, or the nucleotide

sequence that codes for the partial polypeptide sequence that constitutes the higher animal telomerase protein, as represented by sequence no. 2 in the sequence table.

Although DNA and RNA are included in the nucleotides of the present invention, a DNA sequence coding for the entire polypeptide that constitutes rat and human telomerase protein and the DNA sequence that codes for a partial polypeptide sequence that constitutes the human telomerase protein in sequence nos. 1, 13 and 2 respectively in the sequence table are put forth as preferred modes. In addition to the DNA sequences specified by sequence nos. 1, 13 and 2 above pertaining to the nucleotides of the present invention, also included in the nucleotides of the present invention are nucleotides that code for polypeptides that essentially function as partial polypeptides or entire polypeptides of higher animal telomerase proteins in which one or more amino acid residue substitution, insertion and/or deletion has been introduced into the amino acid sequence of the polypeptides coded by these sequences. Changes in amino acid sequences by amino acid residue substitutions, insertions and/or deletions of this type can be carried out by the site-specific mutagenesis technologies described in Nucleic acid Res., Vol. 10, 6487-6500 (1982) and Methods in Enzymol., Vol. 217, 218-227 (1993), and Vol. 217, 270-278 (1993). However, methods are not limited to these, and any method that is usable by a person skilled in the art can also be used.

By using at least part of the higher animal telomerase protein genetic DNA obtained in the manner described above as a hybridization probe or a PCR primer, other types of higher animal telomerase protein genes could be isolated by similar methods. For example, if a PCR primer is used that is derived from a region in which there is the highest degree of homology between the amino acid sequence of rat telomerase protein and tetrahymena telomerase protein (p80), it would then be possible to clarify the amino acid sequence of the human telomerase protein of the corresponding region, and it would also be possible to obtain the entire cDNA thereof.

The higher animal telomerase protein genetic DNA or DNA fragments thereof obtained in the manner described above could then be modified at both ends or at either end, or the DNA or fragment itself could be inserted downstream from a promoter using a known expression vector in accordance with methods that are themselves well known.

The recombinant vector for protein expression produced in his manner could then be introduced into E. coli, yeast, or an animal host, or into other known cells, following a method that is itself well known, thereby producing transformants.

With regard to the details of the method for producing the higher animal telomerase protein of the present invention, an expression vector is used wherein a promoter is present at a location that allows for transcription of the DNA coding for the higher animal telomerase protein obtained in the manner described above.

In order to industrially produce higher animal telomerase protein, a stable host-vector system must be constructed, and a system must be used whereby higher animal telomerase protein having biological activity can be expressed. The higher animal telomerase protein is a comparatively large protein, and refolding is important in obtaining biological activity. In general, when considering refolding, it is advantageous to use animal cells as the host. The higher animal telomerase can be present as a complex composed of numerous protein and RNA subunits, and when purified from a recombinant vector in the form of higher animal telomerase having biological activity, it is desirable for the animal species from which the higher animal telomerase protein is derived to correspond to the animal host from which the host cells are derived. Moreover, after production of the higher animal telomerase protein using E. coli, it goes without saying that it is possible to carry out reconstruction in the form of an active complex with the other structural components *in vitro*.

Examples of animal cells include CHO cells (organism: hamster), COS cells (organism: monkey), NIH3T3 cells (organism: mouse), Rat-1 cells (organism: rat), and VA-13 cells (organism: human). Expression plasmids that can be used with these cells as hosts preferably contain a promoter derived from a virus gene or derived from the SV40 promoter. The higher animal telomerase protein is inserted on the 5' side downstream from the promoter. In addition, in order to increase production quantities of higher animal telomerase protein, 2-3 connected copies of the higher animal telomerase protein gene can also be inserted on the 5' side, and 2-3 copies of a promoter such as the SV40 promoter can be inserted on the 5' side of each of the higher animal telomerase protein genes. It is preferable for a polyadenylation site to be present downstream from the

higher animal telomerase protein gene, and for example, sites derived from SV40 DNA, the β-globulin gene or the metallothionein gene can be used.

This type of expression vector can employ a selection marker during transformation into animal cells such as CHO cells. When a selection marker is used, a DHFR gene that adds methotrexate resistance or a resistance gene that provides neomycin derivative G-418 resistance can be used. It is preferable for the SV40-derived promoter, for example, to be inserted on the 5' side of each of the resistance markers, and for polyadenylation sites to be inserted on the 3' side of each of the resistance genes. When a resistance gene is inserted in an expression vector for the higher animal telomerase protein, it is desirable for the gene to be inserted downstream from the polyadenylation site of the higher animal telomerase protein gene. A transformation selection marker need not be used in the expression vector, and in this case, it is desirable to use vectors that have a transformation selection marker such as pSV2neo, pSV2gpt or pMTVdhfr in conjunction with the expression vector for the higher animal telomerase protein, and to perform double transformations.

An expressed trait resulting from the expression of said selection marker can be used in order to select animal cells that have been transformed with the vector having the transformant selection marker as well as the vector for expressing the aforementioned higher animal telomerase protein. In addition, with the objective of improving expression quantities of the higher animal telomerase protein, the selection marker can be modified and transformation can be repeated into cells that have already been identified as expressing the higher animal telomerase protein. A specific example of a plasmid vector that can be used for an expression vector is pKCR that contains the SV40 early promoter, a splice sequence DNA derived from the sheep β-globulin gene, a polyadenylation site derived from the sheep β -globulin gene, a polyadenylation site derived from the SV40 early region, an origin of replication site from pBR322 and an ampicillin resistance gene (Proc. Natl. Acad. Sci. USA, 78, 1528 (1981)).

Introduction of the expression vector into animal cells generally can be carried out by a transformation method that employs calcium phosphate or cationic lipid as a DNA carrier. Culturing of the transformed animal cells can be carried out by suspension

culturing or adhesion culturing following common methods. A culture medium such as MEM or RPM11640 is used, and culturing is carried out in the presence of appropriate amounts of insulin, dexamethasone, and transferrin in 5-10% serum. Alternately, culturing can be carried out in the absence of serum. Because it is thought that higher animal telomerase protein is present in large quantities in animal cells that express the higher animal telomerase protein, a protein extract obtained from culturing transformants can be used for the separation and purification of the higher animal telomerase protein. The culture supernatant that contains the higher animal telomerase protein that is produced can be purified by various chromatographic procedures, for example, chromatography employing heparin sepharose or blue sepharose.

When E. coli or a microorganism such as Bacillus subtilis is used as a host, the expression vector preferably contains a ribosome binding sequence (SD), the higher animal telomerase protein gene and a gene for controlling the promoter. Examples of promoters include promoters from E. coli or phage-derived promoters, examples of which include the promoter for tryptophan synthesis enzymes (trp), the lactose operon (lac), the λ phage PL or PR, and the T5 phage early gene promoters P25 and P26. In addition, sequences that themselves have been designed and modified such as the pac promoter (Agric. Biol. Chem. 52, 983-988, 1988) can also be used.

Examples of ribosome binding sequences include sequences derived from E. coli, phage, etc., but a sequence having the consensus sequence consisting of a linkage of 4 bases or more of the sequence that is complementary to the 3' terminal region of the 16S ribosomal RNA produced by DNA synthesis can also be used. The transcription termination sequence is not absolutely necessary, but it is preferable for ρ-independent sequences such as a riboprotein terminator or the trp operon terminator to be used.

The order of these sequences that are required for expression on the expression plasmid preferably runs, from the 5' upstream end, as follows: promoter, SD sequence, higher animal telomerase protein gene, transcription termination element. In addition, a method can be employed for increasing the copy number of the transcriptional units on the vector by inserting multiple units of the SD sequence and the higher animal

telomerase protein gene on the expression vector (method described in Japanese Kokai Patent Application No. Hei 1[1989]-95798).

Various types of affinity columns can be used in order readily to recover and purify the higher animal telomerase protein or partial polypeptide thereof that has been expressed in transformants of E. coli, etc. For example, a protein that has a histidine tag consisting of an amino acid sequence consisting of six or more contiguous histidines has the property of binding to a chelate column. Thus, by placing DNA coding for an amino acid sequence consisting of six contiguous histidines downstream from the promoter, and then placing the higher animal telomerase protein gene further downstream, a higher animal telomerase protein or partial polypeptide thereof that includes the histidine tag can be expressed, and the higher animal telomerase protein or partial polypeptide thereof that has been expressed can be readily purified using a chelate column.

In addition, a polypeptide sequence that is specifically cleaved by a protease such as thrombin, TEV protease or factor X can be included between the histidine tag and the polypeptide or partial polypeptide thereof that constitutes the higher animal telomerase protein, so that by treating the polypeptide with the corresponding protease after chelate column purification, the higher animal telomerase protein or partial polypeptide thereof can be recovered in its natural form. After protease cleavage, the material can be separated and purified by HPLC, etc.

In addition, pUAI2 (Japanese Kokai Patent Application No. Hei 1[1989]-95798) or the commercially available pKK233-2 (Pharmacia) can be provided as examples of vectors that can be used as expression vectors. The pGEX series (Pharmacia) can be used as expression vectors that express proteins as fusion proteins with glutathione-S-transferase derived from Japanese Schistosomatoidea, and pProEX-I (Gibco BRL) can be used as a vector whereby purification can be performed using a histidine sequence. Transformation of hosts can be carried out by common methods. In addition, with insect cells, the Maxbac baculovirus expression kit manufactured by Invitrogen can be used following the manual (Maxbac baculovirus expression system manual version 1.4). In this case, it is preferable to change the distance between the polyhedrin promoter and the start codon in order to increase the expression amount.

Culturing of transformants can be carried out following methods that can be used by persons skilled in the art. Appropriate incubation temperatures are 28-42°C. When the lactose operon (lac) promoter is used, it is necessary to perform induction of expression by adding IPTG to a final concentration of about 1 mM at the point when the absorbance of the bacterial culture reaches 0.5 at a wavelength of 600 nm.

The higher animal telomerase protein or partial polypeptide thereof that has been isolated and purified by the methods described above can then be used for immunizing a mammal such as a monkey, sheep, rabbit rat or mouse so that a monoclonal antibody or polyclonal antibody that is specific for the higher animal telomerase protein can be produced. A determination of specificity can be carried out using extract of the gene product or culture liquid from the transformants into which the expression vector containing the higher animal telomerase protein gene has been introduced.

The higher animal telomerase complex can be concentrated and purified from extract of transformants or immortalized cell lines having telomerase activity using an affinity column on which is immobilized monoclonal antibody or polyclonal antibody specific for the higher animal telomerase protein or a partial polypeptide thereof. In addition, a vector that expresses a fusion protein of the higher animal telomerase protein with glutathione-S-transferase, or a tag sequence such as a polyhistidine sequence can be introduced into a eukaryotic immortalized cell line that has telomerase activity, and extract from the resulting transformants can be purified on a column having immobilized ligand that specifically binds the tag sequence, such as glutathione sepharose (Pharmacia) or Nickel NTA Agarose (Qiagen). Higher animal telomerase complexes can be concentrated and purified by this means. Higher animal telomerase complexes that have been obtained by the types of methods described above can be used as active higher animal telomerase for the evaluation of inhibitors, or can be used as a material for analyzing novel constitutive components and for their isolation and purification.

The "two-hybrid" technique can also be employed, and the isolation and identification of genes encoding proteins that physically bind higher animal telomerase protein with high affinity can be carried out using various transformants including yeast. The Match Maker kit manufactured by Clontech can be used for this purpose.

By using antibodies that are specific for the aforementioned higher animal telomerase protein, the degree of expression of the aforementioned gene can be monitored in terms of protein levels, and the expression condition can be monitored at the level of the gene using nucleic acid probes and PCR primers. By means of this type of method, it is possible to detect cancer cells, and to diagnose illnesses that cause fluctuation in telomerase activity, or to diagnose illnesses that accompany fluctuations in telomerase activity. For example after extracting a sample taken from a patient by an appropriate method, a determination can be made by using specific antibodies in an ELISA method or a western blot method, by using nucleic acid probes in a southern or northern blot method, or by using oligonucleotide primers in a PCR method. Consequently, antibodies that can specifically recognize the polypeptide of the present invention or nucleic acid probes that contain nucleotide sequences that can bind complementary to part or all of the nucleotide sequences of the present invention are useful as effective components in reagents for detecting cancer cells, or in compositions for medical use that can be used in diagnosing cancer.

As is shown in the working examples presented below, it is confirmed that the telomerase protein derived from rat includes an inactive polypeptide with a molecular weight of about 240 kDa as determined by SDS polyacrylamide electrophoresis and an active polypeptide with a molecular weight of about 230 kDa as determined by SDS polyacrylamide electrophoresis. The inactive polypeptide with a molecular weight of about 240 kDa is initially expressed, and the presence of a mechanism for converting the protein to the active polypeptide with a molecular weight of about 230 kDa is demonstrated. Consequently, it is evident to the inventors of the present invention that similar inactive and active form polypeptides are present in other higher animals, and that similar mechanisms are also present for converting the polypeptide in its inactive form into the polypeptide in its active form. These molecular species (subunits) are all included within the scope of the present invention.

By measuring the ratio of the aforementioned active form polypeptide and inactive form polypeptide, it is possible to screen substances that act on the telomerase activation mechanism. This screening method typically includes a process involving

measurement of the ratio of the aforementioned active-form polypeptide and inactive-form polypeptide present in tissues or cells from higher animals that have been administered a test substance, or present in tissues or cells of higher animals that have been cultured in the presence of the test substance, and then comparing this ratio with the ratio obtained in the absence of test substance. Measurement of molecular weights is generally carried out by SDS polyacrylamide electrophoresis.

For example, the molecular weight of a subunit of the telomerase protein contained in cells or tissues that have not been in contact with the test substance can be measured by SDS polyacrylamide electrophoresis, and the ratio of the approximately 240 kDa polypeptide and the approximately 230 kDa polypeptide can be determined. Next, the test substance can be administered, or culturing can be carried out in the presence of the test substance, and measurement of the molecular weights of the telomerase protein subunits contained in the tissue or cells in contact with the test substance can then be carried out in like manner in order to determine the ratio of the approximately 240 kDa polypeptide and the approximately 230 kDa polypeptide. If there is an essential increase in the ratio of the approximately 240 kDa polypeptide in the cells or tissues in contact with test substance relative to the ratio in the case where the cells or tissues are not in contact with the test substance, then it can be determined that the test substance inhibits the activation mechanism for telomerase. On the other hand, if there is an increase in the ratio of the approximately 230 kDa protein, then it can be determined that the test substance stimulates activation of telomerase. Substances that are identified in this manner as having an action on the activation mechanism of telomerase are also understood to be within the scope of the invention.

Working Examples

The present invention is described in additional detail below using working examples, but the scope of the present invention is not limited to the examples presented below.

Working Example 1: Obtaining the rat telomerase protein gene

(1) Assay for gene homologous to the gene for tetrahymena telomerase subunit p80

DNA sequences that could code for an amino acid sequence that is homologous to the tetrahymena telomerase subunit p80 amino acid sequence were investigated using the TBLASTN program after accessing the home page of the National Center for Biotechnology Information via the internet. The results showed that a DNA sequence complementary to an mRNA of unknown function derived from rat PC12 cells stored in the Expression Sequence Tag (EST) DNA sequence databank could code for an amino acid sequence (rat cDNA in Table I below) having weak homology with part of the amino acid sequence of p80 (High Score: 94, probability 1.7×10^{-3}). In the table, the amino acids are expressed in single symbol notation, with X denoting the termination codon.

Table I

p80 (N terminal)	AVYIRNEL
Rat cDNA (N terminal)	XASLYARQQL
p80	YIRTTTNYIVAFCVVH
Rat cDNA	NLRDIANIVLAVAALL
p80	KNTQPFIEKYFNKAVL
Rat cDNA	PACRPHVRRYYSAIVH
p80	LPNDLLEVCEFAQVLY
Rat cDNA	LPSDWNQVAEFYQVWY
p80	I (C terminal)
Rat cDNA	L (C terminal)

(2) Obtaining partial fragments of the rat telomerase protein gene

An upstream termination codon was present in the amino acid sequence derived from rat expressing weak homology with the amino acid sequence of p80 obtained in (1),

and moreover, a methionine was not present as the initiation codon downstream therefrom. It was thus unclear as to whether an mRNA that coded for this amino acid sequence was actually present. In addition, it was also unclear whether the protein having homology to p80 could be biosynthesized. However, it was possible that a DNA sequence that was complementary to the DNA sequence stored in the data bank might code for the amino acid sequence, and that the mRNA corresponding to the product of actual transcription was spliced so that its sequence was changed. It was thus investigated whether or not this mRNA was actually present in rat-derived cells.

First, RNA was prepared from Z19 cells derived from rat 3Y1 cells that had been transformed with adenovirus according to the method of Chomczynski (Anal. Biochem., 162, 156-159, 1987). Specifically, 10^8 Z19 cells were homogenized in a guanidine isothiocyanate solution (4 M guanidine isothiocyanate (Wako Pure Chemical), 25 mM sodium citrate (Wako Pure Chemical), 0.1 M 2-mercaptoethanol and 0.5% sodium sarcosyl (Wako Pure Chemical)), and 0.1 volume of 2 M sodium acetate (pH 4.0) was added and mixed. An equivalent volume of a H_2O-saturated phenol (Wako Pure Chemical) and 0.2 volumes of a mixed solution of chloroform (Wako Pure Chemical) and isoamyl alcohol (Wako Pure Chemical) (49:1 volume ratio) were then added, and the solution was mixed vigorously for 10 min, before being centrifuged for 20 min at 10,000 x g to recover the aqueous supernatant layer. An equivalent volume of isopropanol (Wako Pure Chemical) was then admixed with this recovered aqueous layer, and the mixture was chilled for 1 h at -20°C, before being centrifuged for 20 min at 15,000 x g. The resulting precipitate was re-dissolved in guanidine isothiocyanate solution, an equivalent volume of isopropanol was added, and the solution was cooled for 1 h at -20°C. The solution was then centrifuged for 20 min at 15,000 x g to recover the total RNA.

Purification of the RNA was carried out as described below. Specifically, 0.2 mg of total RNA was dissolved in 1 mM EDTA, 20 mM Tris-HCl (pH 7.5), and after heating for 5 min at 70°C, the solution was cooled on ice. 5 M NaCl solution was then added to this solution to produce a final concentration of 0.5 M, and the solution was applied to an oligo-dT cellulose column (type 7, 1 x 1 cm, Pharmacia). After washing the column with

a 20 mM Tris-HCl buffer (pH 7.5) containing 1 mM EDTA and 0.5 M NaCl, the bound fraction was eluted with sterile deionized water to obtain 4 μg of poly-$(A)^+$ RNA.

1 μg of the poly-$(A)^+$ RNA obtained in the manner described above was then used as a template for the synthesis of cDNA. 10 pmol of random hexamer primer and 200 units of MMLV reverse transcriptase ("Super Script", Gibco BRL) were then added to this cDNA, and a 1st strand was synthesized. Next, 1.4 U of RNase H, 40 U of E. coli DNA polymerase and 15 U of E. coli DNA ligase were then added and the 2nd strand was synthesized. After completion of the reaction, phenol/chloroform extraction was performed, and the supernatant aqueous layer was recovered. 5M ammonium acetate solution was then added in an equivalent amount to the recovered aqueous layer, and 2 volumes of ethanol were admixed. The solution was then centrifuged for 10 min at 15,000 x g in order to recover the cDNA by ethanol precipitation.

The cDNA obtained in the manner described above was then used for analysis of the unknown cDNA sequence located further upstream on the 5' side of the region corresponding to the cDNA sequence (sequence no. 3) obtained by the procedure described in (1). The method of Riley et al. was used (Vectorette method, Nucleic Acid Res., 18, 2887-2890). First, 60 ng of cDNA was treated with T4 polymerase in order to produce blunt terminals, and the material was then incubated for 2 h at 37°C with 10 U of restriction enzyme PvuII (Toyobo; the provided buffer was used). The cut DNA was then purified by phenol/chloroform treatment and ethanol precipitation, and 3 pmol of the Vectorette unit shown in Table II below (with vctA and vctB annealed) was ligated using DNA ligase.

Table II

vctA: 5'-AAGGAGAGGACGCTGTCTGTCGAAGGTAAGGAACGGACGA
GAGAAGGGAGAG-3'

vctB: 5'-CTCTCCCTTCTCGAATCGTAACCGTTCGTACGAGAATCGC
TGTCCTCTCCTT-3'

The cDNA having vectorette unit ligated to the blunt end was then used as a template, and a vctG oligonucleotide primer that would hybridize to one strand of the vectorette unit shown in Table III below was used along with the RaPC5' oligonucleotide primer that would hybridize with the cDNA sequence represented by sequence no. 3 in order to perform PCR. The cDNA including the unknown region on the 5' side upstream from the region to which the RaPC5' oligonucleotide primer binds was thus amplified. The amplification reactions were carried out by common methods, and a thermostatic cycle comprising 1 min at 93°C, 1 min at 65°C and 2 min at 72°C was repeated 35 times using a PCR thermal cycler.

Table III

vctG:	5'-CGGTACCGAATCGTAACCGTTCGTACGAGAATCGCT-3'
RaPC5':	5'-CATACCTGGTAGAACTCGGCTA-3'

The PCR product was treated with phenol/chloroform, and was purified by ethanol precipitation. Some of this product was then ligated to the PT7BlueT vector (Pharmacia) using DNA ligase, and transformed recombinant E. coli was selected with ampicillin. Plasmid DNA was then prepared, and the DNA sequence of the inserted PCR product was determined by the Sanger method using the ABI373A sequencer (Applied Biosystems). The results gave a cDNA wherein the base sequence represented by sequence no. 4 in the sequence table had been inserted into the plasmid RaPC53.

The results of analyzing the base sequence of RaPC53 confirmed that the base sequence spanning nucleic acid nos. 1-170 represented by sequence no. 3 in the sequence table, predicted from the complementary strand DNA, corresponded to the actual base sequence spanning nucleic acid nos. 1-244 of sequence no. 4 in the sequence table derived from rat cells. Because the base sequence spanning nucleic acid nos. 163-172 of sequence no. 3 in the sequence table (5'-TCTCTCCTAG-3') corresponded to the consensus sequence 5'-PyPyPyPyPyPyNCAG-3' of the splicing acceptor, this result was concluded to not be an artifact, but rather, was thought to result from RNA editing via splicing. Consequently, the (T) of base no. 170 in sequence no. 3 in the sequence table

actually became an (A) in the sequence of sequence no. 4, so that the termination codon TAG became AAG coding for lysine. Moreover, it was determined that the open reading frame extended farther along the 5' upstream side.

The amino acid sequence of this open reading frame exhibited additionally high homology (High Score: 125, probability 1.6×10^{-18}) in comparison to the rat-derived amino acid sequence that showed homology with the amino acid sequence of tetrahymena p80 predicted in procedure (1) (High Score: 94, probability 1.7×10^{-3}). It was determined that the (A) of no. 312 of sequence no. 3 in the sequence table was changed to a (T), and that the corresponding amino acid thus changed from asparagine (AAC) to isoleucine (ATC).

(3) Obtaining the total length cDNA for rat telomerase protein

First, poly-$(A)^+$ RNA was obtained from rat 3Y1-derived SV-3Y1-C66 cells transformed with SV40 virus according to the same method as in procedure (1), and cDNA was prepared using the cDNA synthesis kit of Stratagene. Although preparation of cDNA was carried out according to the manual, the 1st strand synthesis reaction was carried out adding both random hexamer oligonucleotides and an oligo-dT primer as the primers at final concentrations of 2 μM each.

Next, after adding EcoRI adapters to the ends of the cDNA using DNA ligase, the reaction product was applied to a Sephacryl S-500 column, and the unreacted EcoRI adapters and cDNA of smaller sizes were removed. The cDNA of the flow-through fraction was then recovered by ethanol precipitation, and was digested beforehand with the restriction enzyme EcoRI. The aforementioned cDNA was then linked, using DNA ligase, to λZAP phage DNA that had been subjected to terminal dephosphorylation. The λZAP phage DNA ligated with the cDNA was then packaged into phage particles. The procedure described above was carried out following the included manual using Gigapack Gold III, manufactured by Stratagene. The resulting phage particles were then used to infect the E. coli strain C600hflA following a common method, and the phage particles were amplified and recovered. Approximately 5,000,000 phage clones were obtained by in a single run.

About 1,000,000 phage clones were used to infect the E. coli strain C600hflA according to a common method, and the cells were incubated on NZY agar media using plates. The phage particles were then transferred onto Nylon film, and 2 sets of replicas were produced. After washing and alkali treatment, the RaPC53 obtained in procedure (2) was labeled with ^{32}P and used as a probe, and the phage clones were screened by hybridization with the probe. The results indicated three positive signals, and the phage particles for these signals were recovered. Cloning was then performed by the same method, and the plasmids containing inserted cDNA regions (RET1, RET2 and RET3) were recovered by *in vivo* excision according to the manual provided by Stratagene.

Restriction enzyme site cleavage site maps were produced for plasmids RET1, RET2 and RET3, and it was found that 1.3 kbp, 2.4 kbp and 6.5 kbp cDNA sequences had been inserted, and that the cDNA sequences were present in overlapping locations as shown in Figure 1. Deletion mutants of the cDNA were then produced according to a common method, and the entire RET1 sequence and part of the DNA sequence of RET2 and RET3 were decoded. Upon assembling these DNA sequences into the restriction enzyme cleavage site map, a large open reading frame extending over approximately 4.6 kbp was discovered. Within this sequence was also present the amino acid sequence of RaPC53 that expresses homology with the tetrahymena p80 amino acid sequence obtained in procedure (2) (High Score: 125, probability 1.6×10^{-18}), and additionally high homology was seen with the amino acid sequence of tetrahymena p80 due to homologous sites (high score: 234, probability 1.1×10^{-49}).

However, based on the fact that a terminal codon was not discovered at the C terminal of the open reading frame, and also based on the results of a number of northern analyses of mRNA extracted from rat cells that were positive for telomerase activity, it was concluded that the actual mRNA from which the resulting cDNA had been derived was extremely large, at near 10 kb. An attempt was thus made to read the 3' region of the cDNA. Specifically, a DNA fragment of the region spanning nucleic acid nos. 4083-5216 in the DNA sequence represented by sequence no. 1 in the sequence table, which was near the 3' end of RET3, was used as a ^{32}P labeled probe, and an additional approximately 1,000,000 phage clones were screened. As a result, 13 new positive

signals were found. Among these colonies, plasmids containing inserted cDNA regions were recovered by in vivo excision from 6 of the clones (RETλ01, 07, 08, 09, 10, and 13).

Upon producing restriction enzyme cleavage site maps of the plasmids RET λ 01, RETλ09 and RETλ13, it was found that the various cDNAs were present in overlapping sites with cDNA sequences of 5.0 kbp, 4.9 kbp and 4.9 kbp respectively. Of these sequences, the RETλ13 sequence was renamed "RET7", and cDNA deletion mutants were produced following a common method in order to decode the entire DNA sequence of RET7. By combining the information from the DNA sequences from plasmids RET1, RET2 and RET3, it was discovered that there was a large open reading frame extending over 7890 bp, including the termination codon (sequence no.1 in the sequence table).

(4) Obtaining rat telomerase protein cDNA - obtaining the upstream sequence (5'-RACE method)

No termination codon within the same frame was found on the 5' side upstream from the ATG site that was farthest towards the 5' end of the cDNA sequence obtained in procedure (3), and so an investigation of the mRNA sequence was carried out farther towards the 5' side using the 5'-rapid amplification of cDNA ends (RACE) method.

The 5'-RACE method was carried out according to the manual, using the 5'-RACE kit manufactured by Clontech. 2 μg of poly-$(A)^+$ RNA obtained from SV-3Y1-C66 cells in procedure (3) and 10 pmol of oligonucleotide primer NcEX3' having a DNA sequence homologous to nucleic acid nos. 1493-1515 of sequence no.1 in the sequence table were mixed, and the mixture was cooled after heating. Reverse transcriptase (Superscript, Gibco BRL), substrate nucleotides and buffer were then added, and a reaction was allowed to occur for 1 h at 42°C. EDTA was then added to stop the reaction, whereupon the template RNA was degraded by alkali treatment. Isopropanol precipitation was then carried out to isolate single-chain cDNA. To half the amount of this cDNA was ligated 4 pmol of 5'-RACE anchor primer (5'-P(+) ANC), and the primers were ligated using RNA ligase. The reaction was carried out for 3h at 37°C in the presence of 25% PEG.

Next, reverse transcription priming was performed with NcEX3', and the single-chain cDNA with anchor DNA sequence added to its 3' end was used as template, while using the oligonucleotide primer RACE-PRM homologous to the anchor DNA and the oligonucleotide primer RaPC5' having a DNA sequence complementary to the region of nucleic acid nos. 1039-1056 of sequence no. 1 in the sequence table. DNA amplification was thus carried out by PCR. 1/20 the amount of single-chain cDNA and 10 pmol of each primer were used in the reaction, and PCR was carried out using Taq polymerase manufactured by Gibco BRL according to the included manual. In order to prevent non-specific DNA amplification, the reaction was initiated by the manual hot start method, and 35 repetitions were performed with a cycle comprising 30 sec at 94°C, 1 min at 55°C and 2 min at 72°C.

The PCR product was introduced into the pT7BlueT vector, and the results of reading the DNA sequence of the vectors containing the amplified DNA showed that 10 of these clones had nearly the same DNA sequence. Of these clones, RACE3 and RACE5 that were typical examples of clones were present in the locations shown in Figure 1. It was thus clear that reverse transcription and extension of the cDNA was possible up to 200 base pairs upstream from the 5' ATG of nucleic acid nos. 199-201 in sequence no.1 in the sequence table. A termination codon that matched the frame of sequence no.1 was discovered the 5' region upstream from the ATG site of base nos. 199-201 in sequence no.1 in the sequence table, but the lengths of the amplified DNA were not uniform, and it was concluded that there is a high possibility that cDNA was obtained that corresponded with the 5' end of actual mRNA.

Working Example 2: Obtaining the human telomerase protein gene

(1) Obtaining a partial fragment of the human telomerase protein gene

When the homology between the amino acid sequence of tetrahymena p80 and the rat telomerase protein amino acid sequence obtained in procedure (3) of Working Example 1 was investigated, upon discovering a number of identical amino acid sequences, it was considered possible that these regions were conserved over a wide range of species. Thus, it was expected that the cDNA of the telomerase proteins specific

to various species of animals other than tetrahymena and rat could be recovered by constructing degnenerate PCR primers from the amino acid sequences of these regions, and using these primers in PCR methods.

First, HPET5 (sequence no. 5 in the sequence table) that corresponded to amino acid nos. 379-384 of sequence no.1 in the sequence table was used as the sense primer, and HPET3 (sequence no. 6 in the sequence table) corresponding to amino acid nos. 532-537 of sequence no. 1 in the sequence table was used as the antisense primer. PCR was performed by a common method using, as templates, cDNA derived from rat SV-3Y1-C66 cells obtained in procedure (3) of Working Example 1 and cDNA derived from PA-1 cells from the human ovarian teratoma cell line PA-1 by a similar method. However, the target DNA was not amplified from the cDNA of PA-1 cells or from the cDNA of SV-3Y1-C66 cells used a positive control.

Next, HPET5-2 (sequence no. 7 in the sequence table) corresponding to amino acid nos. 376-385 in sequence no.1 in the sequence table, and HPET5-3 (sequence no. 8 in the sequence table) corresponding to amino acid nos. 380-388 of sequence no. 1 in the sequence table were used as sense primers, and HPET3-2 (sequence no. 9 in the sequence table) corresponding to amino acid nos. 532-540 of sequence no.1 in the sequence table or HPET3-3 (sequence no. 10 in the sequence table) corresponding to amino acid nos. 534-542 of sequence no. 1 in the sequence table were used as antisense primers. PCR was carried out using 4 primer combinations, along with cDNA derived from SV-3Y1-C66 cells and cDNA derived from PA-1 cells as templates.

The PCR product was then subjected to agarose gel electrophoresis, and a UV illuminator was used to observe the gels wherein the DNA was stained with ethidium bromide. The predicted DNA fragment of about 500 bp was amplified in the case where a combination of HPET5-2 or HPET5-3 with HPET3-2 was used with cDNA derived from SV-3Y1-C66 cells as a template. In addition, the approximately 500 bp DNA fragment was similarly amplified when HPET5-2 and HPET3-2 were used as primers in combination with the cDNA derived from PA-1 cells as a template. The DNA sequence of this DNA fragment was then read by subcloning into pT7Blue plasmid, and a DNA sequence was obtained that exhibited about 77% homology at the base level with the

corresponding rat cDNA, and exhibited 76% homology at the amino acid level (Figure 2, sequence no. 2 in the sequence table).

Based on the DNA sequence information that had been obtained, oligonucleotide primers were designed that could amplify a human telomerase protein cDNA fragment by PCR. hTPC5 (sequence no. 11 in the sequence table) corresponding to nucleic acid nos. 92-114 of sequence no. 2 in the sequence table was used as the sense primer, and hTPC3 (sequence no. 12 in the sequence table) corresponding to nucleic acid nos. 433-455 of sequence no. 2 in the sequence table was used as the antisense primer, and PCR was carried out using cDNA derived from the mRNA of various types of human cells as a templates.

First, human placental total RNA, total RNA from the B-cell leukemia cell line Raji, poly-$(A)^+$ RNA from the human squamous cell carcinoma cell line A431, and poly-$(A)^+$ RNA derived from the human breast cancer cell lines BT474, SKBR3, BSMZ and MCF7 were obtained by the method of Chomczynski (Anal. Biochem., 162, 156-159, 1987) using a kit manufactured by Pharmacia, and the cDNA was synthesized using the First Strand Synthesis kit manufactured by Pharmacia.

About 1/20 the amount of these cDNAs were used as templates, and PCR was carried out using hTPC5 and hTPC3 as primers. Amplitaq Gold (Perkin-Elmer) was used as the DNA polymerase, and after treatment for 10 min at 95°C, a thermal cycle comprising 30 sec at 95°C, 30 sec at 65°C and 30 sec at 72°C was repeated 35 times. The results gave amplification of the predicted approximately 390 bp DNA when cDNA derived from human cancer cells was used as a template, but no amplification was detected in the negative control where cDNA derived from the human placental total RNA was used as a template.

This result indicates that human telomerase protein cDNA fragment can be amplified when hTPC5 and hTPC3 are used as primers. PCR was then carried out by a similar method on 100,000 phage from the human placental cDNA library manufactured by Clontech, but DNA amplification was not observed. However, when 1,000,000 phage were used as templates, DNA of the predicted size was amplified.

Then, using the aforementioned cDNA library as a vector, two oligonucleotide primers corresponding to the 3' side and 5' side of the cDNA insertion site of λgt10 (5'λgt10 and 3'λgt10 respectively, manufactured by Clontech) were used as primers along with hTPC5 and hTPC3, and an attempt was made to obtain a cDNA fragment of the unknown region downstream from the 3' end of hTPC3 or upstream from the 5' end of hTPC5. 1,000,000 phage from the cDNA library were used as templates, and PCR was carried out by the method described above using four primer combinations (hTPC5 with 5'λgt10 or 3'λgt10 and hTPC3 with 5'λgt10 or 3'λgt10). 55°C was used as the annealing temperature rather than 65°C. The results indicated that a DNA fragment was amplified that corresponding to an approximately 1.5 kbp region upstream on the 5' side of hTPC5.

(2) Obtaining total cDNA from the human telomerase protein gene

First, the total RNA was obtained from about 100,000,000 each of Raji cells and PA-1 cells following the method of Chomczynski (Anal. Biochem., 162, 156-159, 1987) using the RNAzol solution (Tel-Test). The resulting total RNA was then applied to an Oligo-dT cellulose column (type 7, 1 x 1 cm, Pharmacia) to obtain about 100 μg of poly-$(A)^+$RNA.

5 μg of poly-$(A)^+$ RNA was used as template for the synthesis of cDNA. In the reaction, reverse transcriptase, ribonuclease H and E. coli DNA polymerase provided in the cDNA synthesis module (Amersham) were used, and double-stranded cDNA synthesis was performed according to the supplied description. Next, T4 DNA polymerase provided in the cDNA synthesis module (Amersham) was used, and blunting of the cDNA terminals was performed. After the reaction, phenol/chloroform extraction was carried out and the aqueous supernatant layer was recovered. 5 M ammonium acetate solution was then added in a volume equivalent to that of the aqueous layer, whereupon 2 volumes of ethanol was admixed. The solution was then centrifuged for 10 min at 15,000 x g, and the cDNA was recovered by ethanol precipitation. The recovered cDNA was dried and then dissolved in 20 μL of sterile deionized water, before being separated into two 10 μL amounts of cDNA (about 2 μg). EcoRI adapters (Takara Shuzo)

were then added to the terminals. Specifically, the material was incubated for 2 h at 16°C with 350 U of T4 DNA ligase (Takara Shuzo) and 20 μL of T4 DNA ligase reaction solution (66 mM Tris-HCl buffer (pH 7.6), 6.6 mM $MgCl_2$ (Wako Pure Chemical), 10 mM dithiothreitol (DTT, Wako Pure Chemical) and 66 mM adenosine 5'-triphosphate (ATP, Sigma)). 200 pmol of EcoRI adapter was thereby ligated to the terminals of the cDNA.

The reaction product was then applied to a Sephacryl S-200 column (1 x 4 cm), and 10 mM Tris-HCl buffer (pH 7.5) containing 0.5 mM NaCl and 1 mM EDTA was then used to elute the cDNA with EcoRI adapter attached to its terminals. The eluted cDNA was then recovered by ethanol precipitation, and after drying the precipitate, the DNA was dissolved in 2 μL of sterile deionized water. After pre-digesting with restriction enzyme EcoRI (Takara Shuzo), 1 μg of λZAP phage DNA (Stratagene) that had been subjected to terminal dephosphorylation and the aforementioned cDNA with EcoRI adapters attached (400 ng) were incubated for 18 h at 16°C in T4 DNA ligase reaction solution (5 μL) in order to ligate. The λZAP phage DNA ligated to the cDNA was then packaged into phage particles using Gigapack II Gold (Stratagene).

The resulting phage particles were then used to infect E. coli strain C600hflA, and were amplified according to a common method before recovering the phage particles. In a single procedure, about 2,000,000 phage clones were obtained per 100 ng of cDNA. About 1,000,000 phage clones were used to infect E. coli strain C600hflA following a common method, and the cells were incubated on plates using NZY agar medium. The phage particles were then replicated onto nylon film, producing two replica sheets. After washing and alkali treatment, the phage clones were then screened by hybridization with a probe using the probe obtained in procedure (1) of Working Example 2 produced by labeling the human telomerase protein cDNA fragment with ^{32}P. The phage particles were recovered based on the resulting positive signals, and after cloning by the same method, the plasmids containing the inserted cDNA region were recovered by in vivo excision following the manual of Stratagene.

(3) Obtaining the downstream sequence from the 3' end of the entire human telomerase protein cDNA (3'-RACE)

The mRNA obtained in procedure (2) above was used, and the cDNA was amplified using the RACE method employing the Marathon™ cDNA Amplification Kit (Clontech). In the reaction described below, the synthetic DNA primers other than the primers included in the Marathon™ cDNA Amplification Kit were synthesized using the ABI394 DNA Synthesizer. The reaction was carried out using dNTPs and the buffer solution provided in the Marathon™ cDNA Amplification Kit.

First, cDNA was synthesized. 1 μg of purified poly-$(A)^+$ RNA and the cDNA reverse transcription primer 5'-TTCTAGAATTCAGCGGCCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTT(G/A/C)(G/A/C/T)-3' (52 nucleotides) were treated at 37°C with reverse transcriptase to obtain single chain cDNA. The second chain extension reaction and terminal blunting were then performed, and adapter primers (5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3' (44 nucleotides) and 5'-PO_4-ACCTGCCC-NH_2-3' (8 nucleotides)) were ligated to both terminals of the cDNA. 10 μL of the reaction liquid obtained upon completion of the reaction was diluted to produce 50 μL, and 1 μL amounts were used in subsequent amplification reactions.

The amplification reaction was carried out using a primer complementary to the adapter primer attached to the 3' end (5'-CCATCCTAATACGACTCACTATAGGGC-3' (27 nucleotides) and a primer complementary to part of the cDNA sequence of human telomerase protein. The amplification was carried out using Taq DNA polymerase. The entire quantity of reaction solution was adjusted to 50 μL, and after incubating for 1 min at 94°C, 30 cycles were carried out comprising incubation for 30 sec at 94°C, 30 sec at 60°C and 5 min at 68°C. The reaction was completed by a final incubation for 7 min at 72°C. 1/10 of the reaction solution was analyzed by 5% PAGE. In addition, 5 μL of the aforementioned reaction solution was diluted 50x, and 5 μL of this solution was used in a second amplification reaction.

The second amplification reaction was carried out in the same manner as the first amplification reaction. 5 μL of diluted reaction solution was used as template, and a primer located inside of the primer used for the first amplification reaction that was complementary to part of the cDNA sequence of the human telomerase protein was used along with 5'-ACTCACTATAGGGCTCGAGCGGC-3' (23 nucleotides) in an amplification reaction using Taq DNA polymerase. The entire volume of reaction liquid was adjusted to 50 μL, and after incubating for 1 min at 94°C, 30 cycles comprising incubation for 30 sec at 94°C, 30 sec at 60°C and 5 min at 68°C were carried out. The reaction was completed by a final incubation for 7 min at 72°C. After completion of the reaction, 1/10 of the reaction solution was analyzed by 5% PAGE.

Next, the amplified cDNA fragment was recovered from among the gel fragments and was purified, before being inserted into the cloning region of the plasmid vector pCRII (Invitrogen). E. coli strain JM109 was transformed with this recombinant vector, and X-Gal-IPTG-LB-Amp agar plates were used in order to find 3 transformants that were resistant but were not colored by X-Gal. These transformants were analyzed by plasmid DNA preparations following a common method. The prepared plasmid DNA was then used in determining the cDNA base sequences. The results gave a cDNA fragment having a base sequence in the 3' untranslated region.

(4) Obtaining the region upstream from the 5' end of the entire human telomerase protein cDNA (5'-RACE method)

The reactions of the 5'-RACE method were carried out based on those of the 3'-RACE method. The synthetic DNA primers other than those supplied with the Marathon™ cDNA Amplification Kit were synthesized using the ABI394 DNA synthesizer. The reaction was carried out using the dNTPs and buffer solution provided with the Marathon™ cDNA Amplification Kit. As in the 3'-RACE reactions, cDNA with adapter primers attached to both ends was used as the template. The first amplification reaction was carried out using a primer complementary to part of the cDNA sequence of human telomerase protein and the primer 5'-CCATCCTAATACGACTCACTATAGGGC-3' (27 nucleotides) that was used in the

reaction of the 3'-RACE method complementary to the adapter primer attached to the 3' terminal. The total volume of the reaction solution was adjusted to 50 μL ,and the amplification reaction was carried out using Taq DNA polymerase. The reaction was carried out by incubation for 1 min at 94°C, followed by 30 cycles comprising 30 sec at 94°C, 30 sec at 60°C and 5 min at 68°C, and the reaction was completed by a final incubation for 7 min at 72°C. After the reaction, 1/10 of the reaction solution was analyzed by 5% PAGE. 5 μL of the aforementioned reaction solution was also diluted 50x, and 5 μL of this solution was used as a template in the second amplification reaction.

The second amplification reaction was carried out based on the first amplification reaction. The reaction was carried out using, as primers, 5'-ACTCACTATAGGGCTCGAGCGGC-3' (23 nucleotides) and a primer located inside the primer used in the first amplification reaction that was complementary to part of the cDNA sequence of the human telomerase protein. After incubating for 1 min at 94°C, the reaction was carried out in 30 repetitions of a cycle comprising incubation for 30 sec at 94°C, 30 sec at 60°C and 5 min at 68°C, and the reaction was completed by a final incubation for 7 min at 72°C. After the reaction, 1/10 of the reaction solution was analyzed by 5% PAGE. The cDNA amplified from the gel fragment was recovered and purified, and after inserting into the cloning region of the plasmid vector pCRII, E. coli strain JM109 was transformed using the resulting recombinant vector. X-Gal-IPTG-LB-Amp agar plates were then used in order to find 3 transformants that were resistant but were not colored by X-Gal. Plasmid DNA was then prepared following a common method from these three transformants, analyses were carried out using the prepared plasmid DNA, and the base sequences were determined. The results gave a cDNA fragment having a base sequence in the 5' untranslated region of human telomerase protein.

Working Example 3: Obtaining the human telomerase protein gene

(1) Obtaining the entire cDNA for the human telomerase protein gene

First, a cDNA library was constructed using PA-1 cells in the same manner as when the rat telomerase protein gene was obtained. This library was screened using a

probe that was a PCR product formed using hTPC5 (sequence no. 11 in the sequence table) described above and hTPC3 (sequence no. 12 in the sequence table) described above as primers. The entire cDNA for the human telomerase protein gene was thus obtained.

First, poly-$(A)^+$ RNA was obtained from PA-1 cells. Specifically, 10^8 cells were homogenized in a guanidine isothiocyanate solution, and 0.1 volume of 2 M sodium acetate (pH 4.0) was added and mixed. An equivalent volume of H_2O-saturated phenol and 0.2 volume of chloroform/isoamyl alcohol mixture were then added to the homogenate, and were mixed vigorously, whereupon the solution was centrifuged to recover the aqueous supernatant layer. Isopropyl alcohol was then admixed in an equal volume with the recovered aqueous layer, and after chilling for 1 h at -20°C, the solution was centrifuged. The resulting precipitate was dissolved again in guanidine isothiocyanate solution, an equivalent volume of isopropanol was added, and after chilling for 1 h at -20°C, the total RNA was recovered by centrifuging.

The total RNA was dissolved in 1 mM EDTA, 20 mM Tris-HCl (pH 7.5), and after heating for 5 min at 70°C, the solution was chilled on ice. NaCl solution was then added to this solution to produce a final concentration of 0.5 M, and the solution was applied to an oligo-dT cellulose column (type 7, 1 x 1 cm, Pharmacia). After washing the column with 20 mM Tris-HCl buffer (pH 7.5) containing 1 mM EDTA and 0.5 M NaCl, the bound fraction was eluted with sterile distilled water to obtain poly-$(A)^+$ RNA.

This poly-$(A)^+$ RNA was then used for the production of cDNA employing a DNA synthesis kit from Stratagene. 1st strand synthesis was carried out by adding, as primers, random hexamer oligonucleotides and an oligo-dT primer, both at final concentrations of 2 μM. The cDNA terminals were then blunted using T4 DNA polymerase, and EcoRI adapters were added to the terminals. The reaction product was then applied to a Sephacryl S-500 column, and the unreacted EcoRI adapters and the small-size cDNA were removed. The cDNA was then recovered by ethanol precipitation, and was inserted into λZAP phage DNA.

The λZAP phage DNA ligated with the cDNA was then packaged into phage particles using Gigapack Gold III from Stratagene. About 10,000,000 phage clones were obtained from a single preparation.

About 1,000,000 phage clones were used to infect E. coli strain C600 hflA following a common method, and the bacteria was then cultured on plates using NZY medium. The phage particles were then transferred onto nylon film, producing 2 replicas, whereupon the sheets were washed and alkali treated. A PCR product produced using hTPC5 and hTPC3 as primers was then labeled with ^{32}P and was used as a probe for screening the phage colonies by hybridization with the probe. The results gave two positive signals, from which the phage particles were recovered. Cloning was then carried out using a similar method, and the plasmids (pHB01 and pHB04) that contained cDNA inserts were recovered by *in vivo* excision.

Upon producing restriction enzyme cleavage site maps for pHB01 and pHB04, it was found that cDNA with sizes of 1.1 kbp and 7.4 kbp respectively had been inserted, and as shown in Figure 4, it was found that the cDNA fragments were in overlapping locations. Deletion-mutated cDNA was then produced according to a common method, and the DNA sequence of pHB01 and pHB04 was read. This DNA sequence was found to cover a region spanning about 8.1 kbp, as determined by combining the restriction enzyme cleavage maps. In this region, a long open reading frame including a stop codon on the C-terminal side was found. The amino acid sequence predicted from this open reading frame showed high homology, with an identity of over 70%, with respect to the amino acid sequence of the C-terminal side of rat telomerase protein. The sequence was thus judged to be the sequence for human telomerase protein.

(2) Obtaining the human telomerase protein cDNA and the upstream sequence (5'-RACE method)

The DNA sequence obtained in procedure (1) was the sequence after nucleic acid no. 756 in the DNA sequence shown in Sequence no. 13 in the sequence table, but from a comparison of the primary structure of rat telomerase protein, it was thought that the open reading frame would extend farther towards the N-terminal side. Consequently, the 5'

side of the mRNA sequence was investigated using the 5'-Rapid Amplification of cDNA Ends (RACE) method.

The 5'-RACE method was carried out according to the manual using the 5'-RACE kit manufactured by Clontech. 2 μg of poly $(A)^+$ RNA obtained from the PA-1 cells in procedure (1) was mixed with 10 pmol of oligonucleotide primer TLPCM3 having a DNA sequence complementary to the region spanning nucleic acid nos. 1165-1187 of sequence no. 13 in the sequence table, and the mixture was heated and allowed to cool. Reverse transcriptase (SuperScript II, Gibco BRL), substrate nucleotides and buffer solution were added to the reaction mixture, and a reaction was allowed to occur for 1 h at 42°C. EDTA was then added to stop the reaction, and the template RNA was degraded by alkali treatment. The single-chain cDNA was then isolated by isopropanol precipitation. In addition, to half of this cDNA was ligated 4 pmol of 5'-RACE anchor primer (5'-P(+)ANC) using RNA ligase [sic].

Next, reverse transcription priming was carried out with TLPCM3, and using the single-chain cDNA with anchor DNA sequence attached to the 3' end as a template, DNA amplification was carried out by PCR using the oligonucleotide primer TLPNE having a DNA sequence that is complementary to the region spanning nucleic acid nos. 1024-1046 of sequence no. 13 in the sequence table along with the oligonucleotide primer RACE-PRM2 that is complementary to the anchor DNA. In the reaction, 1/20 the amount of single-chain cDNA and 10 pmol of each primer were used, and PCR was carried out according to the supplied manual using Taq polymerase manufactured by GIBCO BRL. In order to prevent non-specific DNA amplification, the reaction was initiated by the hot start method as described in the manual, and 35 repetitions were carried out using a cycle comprising 30 sec at 94°C, 1 min at 60°C and 2 min at 72°C.

The PCR product was inserted into the pT7BlueT vector, and the DNA sequences of the vectors containing the amplified DNA insert were read. Of these inserts, 3 clones had nearly the same DNA sequence. Of these clones, RACE-L4, a typical clone, was present in the location shown in Figure 4. An initiation codon occurs at nucleic acid nos. 156-158 of sequence no. 13 in the sequence table, and a stop codon occurs in the same frame at nucleic acid nos. 144-146 upstream in the same sequence. The length of

amplified DNA was nearly identical up to 157 bp upstream on the 5' side of the initiation codon, and thus it was concluded that there was a high probability that cDNA was obtained that corresponding to the 5' end of actual mRNA.

Working Example 4: Obtaining recombinant rat telomerase protein and production of a specific antibody

A fusion protein (GST-p80hom) formed from Japanese Schistosomatoidea glutathione-S-transferase and rat telomerase protein (partial polypeptide corresponding to amino acid nos. 217-345 of sequence no. 1 in the sequence table) was expressed in E. coli, and the purified gene product was used for immunizing rabbits. Next, the same region of the rat telomerase protein was expressed as a fusion protein (6His-p80hom) with a histidine hexamer using another expression vector, and the purified gene product was used for producing an affinity column. Polyclonal antibodies specific for rat telomerase protein were thus obtained from the rabbit antiserum (polyclonal antibody specific for the region corresponding to amino acid nos. 217-345 of sequence no.1 in the sequence table).

After cutting the expression plasmid vector pGEX2T (Pharmacia) with the restriction enzyme Sma I, an oligonucleotide linker having a Hind III cleavage site was inserted to produce the expresion vector pGEXH12. This vector was cleaved with the restriction enzyme EcoRI, and T4 polymerase (Toyobo) was used in order to blunt its ends. The vector was then cleaved with the restriction enzyme Hind III. Next, the plasmid RaPC53 containing the rat telomerase protein cDNA fragment was cut with the restriction enzyme Bam HI, and the terminals were blunted using T4 polymerase (Toyobo). The vector was then cleaved with the restriction enzyme Hind III, and the partial DNA fragment of the rat telomerase protein cDNA was isolated by polyacrylamide gel electrophoresis (Hind III-Bam HI DNA fragment with blunt ends spanning about 390 bp corresponding to nucleic acid nos. 648-1034 of sequence no. 1 in the sequence table). The Hind III blunt end pGEXH12 vector and the DNA fragment derived from the rat telomerase protein cDNA obtained in the manner described above were ligated using a DNA ligation kit (Takara Shuzo), and the resulting recombinant

vector was used in the transformation of E. coli strain JM109 (Toyobo). A restriction enzyme map for each of the plasmids from the ampicillin resistant clones was produced, and the correctly recombined plasmids were retained and stored as pGEXp80hom/JM109.

pGEXp80hom/JM109 was then used to inoculate 50 mL of LB medium containing ampicillin, and the medium was shaking cultured overnight at 37°C. The following day, the culture was diluted 10x with the same medium and was cultured for 1 h at 37°C, whereupon IPTG was added to a final concentration of 0.3 mM. GST-p80hom was expressed having a molecular weight of about 44 kDa by SDS PAGE. Recombinant E. coli expressing GST-p80hom was then suspended in buffer containing sodium sarcosyl at a final concentration of 1.5% following the method of Frangoni (Anal. Biochem. 210, 179, 1993), and Triton X-100 was added to a final concentration of 2%. Glutathione sepharose beads (Pharmacia) were then added and suspended. After maintaining for 40 min at a constant temperature of 4°C while suspending, the beads were washed with phosphate buffer (PBS) containing 1% Triton X-100, and the beads were packed into a column. The GST-p80Hom bound to the beads was then eluted with Hepes buffer containing 0.1% Triton X-100 and 25 mM reduced glutathione.

Typically, 0.7 mg of GST-p80hom was obtained from 100 mL of recombinant bacterial culture. The GST-p80hom was then treated with thrombin and the fusion protein was cut into two pieces, an approximately 29 kDa GST piece and an approximately 16 kDa rat telomerase protein piece (region corresponding to amino acid nos. 217-345 in the rat telomerase protein expressed by sequence no.1 in the sequence table) as determined by SDS PAGE. The latter piece was subjected to N-terminal amino acid sequencing by the Edman method after immobilizing the protein on PVDF film, and the protein was found to be identical to the predicted amino acid sequence. Two adult male rabbits of the traditional Japanese species (R1 and R2) with body weights of about 2.6 kg were inoculated with a mixture of GST-p80hom and Freund's adjuvant in the amount of 100 μg /dose/animal, and antiserum was obtained.

In order to construct an affinity column for purifying antibody specific for rat telomerase protein from the aforementioned antiserum, an antigen of the same region was expressed using a histidine hexamer tag sequence instead of GST, and the protein was

purified in a similar manner. First, plasmid RaPC53 was cut with the restriction enzymes Hind III and Bam HI to produce an approximately 390 bp Hind III-Bam HI DNA fragment of rat telomerase protein cDNA (corresponding to nucleic acid nos. 648-1034 of sequence no. 1 in the sequence table). This fragment was isolated, and was sub-cloned into the Hind III-Bam HI site of pBlueScript (Toyobo). Using the restriction enzymes Xho I and Not I, a Xho I-Not I DNA fragment containing the DNA fragment corresponding to nucleic acid nos. 648-1034 of the rat telomerase protein cDNA (sequence no. 1 in the sequence table) was isolated from this plasmid. A DNA ligation kit (Takara Shuzo) was then used in order to ligate this fragment into the expression plasmid vector pProEX-1 (Gibco BRL) that had been cleaved with the restriction enzymes Sal I and Not I. The resulting recombinant vector was then used to transform E. coli strain JM109 (Toyobo). Restriction enzyme cleavage site maps of the resulting plasmids obtained from each of the ampicillin resistant clones were produced, and correctly recombined plasmids were retained and stored as pProEXp80hom/JM109.

pProEXp80hom/JM109 was then used to inoculate 50 mL of LB medium containing ampicillin, and the medium was shaking cultured overnight at 37°C. The following day, the culture was diluted 10x with the same medium and was cultured for 1 h at 37°C, whereupon IPTG was added to a final concentration of 1 mM. By this means, 6His-p80hom with a molecular weight of about 18 kDa as determined by SDS PAGE was expressed. The recombinant E. coli that was used for expressing the 6His-p80hom was then suspended in binding buffer containing 6 M guanidine hydrochloride following the protocol of Qiagen, and the solution was added to Ni-NTA-Agarose (Qiagen). After washing the beads, the bound 6His-p80hom was eluted with Tris/phosphate buffer at pH 4.3 containing 6 M urea, and after neutralizing the fraction containing the purified 6His-p80hom, the urea was diluted by dialysis against PBS. The insoluble matter was then removed by centrifuging, and Affigel 10 (BioRad) was suspended in the supernatant. Affinity beads were thus produced that were cross-linked with 6His-p80hom. Typically, 0.7 mg of soluble 6His-p80hom was obtained from 100 mL of pProEXp80hom/JM109 bacterial culture, and 95% or greater of the protein is crosslinked to Affigel 10.

Following the method described in "Antibody" (Ed Harlow et al., Ed. Cold Spring Harbor Laboratory Press), 175 μg of antibody (R1-41d) was obtained from 2 mL of hyperimmunized serum taken on the 7th week from R1 that had been immunized with GST-p80, and 86 μg of antibody (R2-41d) was obtained from 2 mL of hyperimmunized serum taken on the 7th week from R2 that had been similarly immunized. It was confirmed by western blotting that these purified antibodies did not react to GST, and reacted only with rat telomerase protein (region corresponding to amino acid nos. 217-345 of the rat telomerase protein expressed by sequence no. 1 in the sequence table).

Working Example 5: Evaluation of antibody specific for rat telomerase protein by immunoprecipitation and telomerase activity measurement.

It was proven in the manner described below that the rat or human telomerase protein cDNA obtained in Working Example 1 or Working Example 3 actually coded for rat or human telomerase protein. Specifically, antibody specific for the recombinant rat telomerase protein fragment obtained in Working Example 4 was used in order to investigate whether telomerase activity in rat or human cell extracts could be immunoprecipitated.

First, the total IgG (PI-1) was purified from pre-immunized serum of R1 using protein A sepharose (Pharmacia), and protein A sepharose was pre-coated with this IgG along with purified IgG obtained from the hyperimmunized serum of R1, R1-41d (derived from serum 7 weeks after initial immunization) and R1-116d (derived from serum 16 weeks after initial immunization), for a total of 3 types of IgG. S100 extracts were then prepared following the method of Counter et al. (EMBO J., 11, 921, 1992) from the rat liver cancer cell line AH66F and the ovarian teratoma cell line PA-1. An equivalent volume of 1% CHAPS/1x Hypo buffer solution (Counter et al., ibid.) was then added to these extracts to produce mixtures, and aforementioned protein A sepharose beads coated with 5 μg of IgG was added to 150 μL of these solutions. The mixtures were then maintained for 1.5 h at a constant temperature of 4°C. The respective beads were then washed with 0.5% SHAPS/1xHypo buffer, and the beads were then suspended in telomerase reaction solution in order to measure the telomerase activity.

Telomerase activity was measured according to the method of Tatematsu et al. (Oncogene, 13, 2265-2274, 1996) using an ELISA method that employed anti-digoxigenin antibody and digoxigenin-labeled dUTP. Specifically, biotin labeled oligonucleotide bpTG3 (biotinylated 5'-GTAAAACGACGGCCAGTTTGGGGTTGGGGTTGGGGT TG-3') was used as a primer for the extension reaction of telomerase, and 0.8 mM of each monodeoxynucleotide (TTP, dATP, dGTP) were used as substrates in a 1 h reaction at 30°C. The enzymatic reaction was stopped by the addition of an excess of EDTA.

Meanwhile, streptavidin (Gibco BRL) was crosslinked to a polycarbonate 96-well microtiter plate (Takara) using EDC (Sigma), and blocking was carried out for 2 h at 37°C using a blocking agent (Boehringer-Mannheim Yamanouchi). 25 μL of the telomerase extension reaction product diluted with TBS was added to each well, and streptavidin binding was allowed to occur on the plate while maintaining a constant temperature of 37°C for 30 min. The sample solutions were then removed, and an excess of biotin solution was added. The solution was maintained for 30 min at 37°C, thereby blocking the excess streptavidin.

Each well was then washed, and a PCR reaction solution was added that contained 20 mM Tris-HCl (pH 8.3), 75 mM KCl, 0.05% W-1, 1.5 mM $MgCl_2$, 4 μM bpTG3, 1 μM oligonucleotide primer pTAG gamma (5'-CAGGAAACAGCTATGACCCCTA ACCCTAACCCTAACCCT-3'), 50 μM each of dATP, dCTP, dGTP, 25 μM TTP, 1 μM digoxigenin-dUTP (Boehringer-Mannheim Yamanouchi) and 1 U of Taq polymerase (Gibco BRL) treated with Taq start antibody (Toyobo). PCR amplification was carried out using the Takara PCR Thermal Cycler (34 cycles comprising 30 sec at 93°C, 30 sec at 69°C and 1 min at 72°C).

Streptavidin prepared at 5 mg/mL using 50 mM sodium carbonate buffer (pH 9.6) was introduced at 100 μL/well into a 96-well microtiter plate composed of white polystyrene, and the streptavidin was allowed to coat the plate by maintaining the temperature at 37°C for 1 h. The streptavidin solution was then removed, blocking buffer was introduced at 150 μL/well, and blocking was carried out for 2 h at 37°C. PCR product diluted 20x with TBS was then added to the wells in amounts of 100 μL/well, and the

solution was maintained at a constant temperature of 37°C for 30 min in order to cause binding to the plate. Each well was then washed 5 times at 150 μL/well with 0.05% Tween 20/TBS, and alkali phosphatase-labeled digoxigenin antibody (Boehringer Mannheim Yamanouchi) diluted 5000x with TBS was added. The plate was then maintained at a constant temperature of 37°C for 30 min, and the plate was washed 5 times at 150 μL/well with 0.05% Tween 20/TBS. CSPD (disodium 3-(4-methoxyspiro(1,2-dioxetane-3,2-(5-chloro)tricyclo(3,3,1,13,7)decan)-4-yl)phenyl phosphate) (Tropix) diluted 100x with 0.1 M diethanolamine buffer (pH 9.5) was then added, and photoluminescence was allowed to occur for 30 min at room temperature, while measuring emissions using a luminometer (Berthold Japan).

As shown in Figure 3, the results indicated that in both experiments where rat cancer cell extract and human cancer cell extract were used, almost no telomerase activity was seen with beads coated with pre-immunization serum IgG (PI-1) and beads not coated with IgG, whereas high telomerase activity was seen with beads coated with either of the two lots of antibodies specific for the recombinant rat telomerase protein fragment obtained in Working Example 4.

Working Example 6: Evaluation of antibody specific for rat telomerase protein by immunoprecipitation of ^{35}S-methionine labeled rat cancer cell extract

5,000,000 cells of the rat hepatoma cell line AH66F were washed with methionine-deficient Dulbecco's modified MEM (DMEM) containing 10% dialyzed fetal calf serum (dFCS), and were then incubated in the same medium to which ^{35}S-methionine had been added. ^{35}S labeling was thus performed, whereupon extraction was carried out using the 0.5% CHAPS/1x Hypo buffer solution used in Working Example 5. Protein A sepharose beads pre-coated with serum IgG from pre-immunized rabbit R1 or serum IgG from rabbit hyper-immunized with recombinant rat telomerase protein fragment were then added to the extract in amounts based on similar cell numbers, and a constant temperature of 4°C was maintained for 2 h. After washing, thermal denaturation was carried out with Laemmli SDS denaturing buffer, and the samples were run using 6% SDS PAGE. The gels were then fixed with acetic acid, and were treated with Enhance

(NEN). After drying, the gels were subjected to fluorography, and the results showed a clear band at about 300 kDa only for the sample treated with IgG from hyper-immunized serum.

Working Example 7: Expression of human telomerase protein mRNA in human cancer cells and normal tissue

The expression of human telomerase protein mRNA was investigated in human cancer cells and normal tissue using the Multiple Tissue Northern Blot and Human Cancer Cell Line Multiple Tissue Northern Blot products manufactured by Clontech. The human telomerase protein gene cDNA fragment (sequence no. 2 in the sequence table) obtained in procedure (1) of Working Example 2 labeled with ^{32}P was used as a probe, and hybridization was carried out overnight at 42°C in the presence of 50% formamide. Each of the blotting films was subjected to autoradiography after washing with 1x and 0.1x SSPE buffer solution containing 0.1% SDS.

The results gave a detection of clear 10.7 kb bands for the poly-$(A)^+$ RNA derived from human normal tissues such as spleen, thymus, pancreas, testes, ovaries, small intestine, large intestine, heart, uterus, lung, liver, bone marrow and kidney. In addition, RNA blotting gave a short molecule of 8.6 kb in addition to the 10.7 kb for poly $(A)^+$ RNA derived from a human cancer cell line.

Working Example 8: Purification of rat telomerase protein and identification of molecular species

An S100 extract was prepared following the method of Counter et al. (EMBO. J, 11, 1921, 1995) from 3 x 10^9 cells of the rat hepatoma cell line AH66F. This extract was then applied to a heparin sepharose CL-6B column (Pharmacia) equilibrated with TMG buffer (10 mM Tris-HCl, pH 8.0, 1 mM magnesium chloride, 1 mM dithiothreitol, 10% glycerol), and step-wise elution was carried out using potassium chloride. The telomerase activity from the fractions of each elution were measured by the method described in Working Example 5, and the fractions containing activity were collected. These fractions were then applied to a hydroxyapatite column (BioRad) saturated [sic]

with TMG buffer solution containing 50 mM potassium chloride, and after washing with 5 mM KP buffer solution (0.25 mM potassium dihydrogen phosphate, 4.75 mM potassium monohydrogen phosphate, 50 mM potassium chloride, 1 mM magnesium chloride, 1 mM dithiothreitol, 10% glycerol), step-wise elution was carried out using 0.5 M KP buffer solution (25 mM potassium dihydrogen phosphate, 475 mM potassium monohydrogen phosphate, 50 mM potassium chloride, 1 mM magnesium chloride, 1 mM dithiothreitol, 10% glycerol).

The fractions having telomerase activity were collected, and were applied to an anion exchange column (product name, RESOURCE Q, Pharmacia) saturated [sic] with TMG buffer solution containing 50 mM potassium chloride, and step-wise elution was performed using potassium chloride. Next, the fractions having telomerase activity were collected, and were applied to a metal (Zn^{2+}) chelate affinity column (product name HiTrap Chelating, Pharmacia) saturated [sic] with TMG buffer solution containing 0.5 M potassium chloride and 1 mM imidazole (not containing dithiothreitol), whereupon step-wise elution was carried out using imidazole. The eluted fractions having telomerase activity were then subjected to centrifugal separation on a 15-40% glycerol concentration gradient (Beckman SW28 rotor, 25000 rpm, 2°C, 24 h). The results gave a protein with a sedimentation coefficient of 44S for the protein corresponding to the telomerase activity, and the molecular weight of this protein was calculated at about 1500 kDa.

In addition, each of the fractions produced by centrifugal separation on the glycerol concentration gradient were isolated by 6% SDS PAGE, and upon western blotting with an antibody specific for recombinant rat telomerase protein obtained in Working Example 4, three bands for antibody reactivity were seen in the protein fractions exhibiting telomerase activity (molecular weights of about 240 kDa, 230 kDa and 55 kDa as determined by SDS PAGE). Of these proteins, the 55 kDa band was confirmed, on the basis of thermal treatment experiments, to be a protein degradation product of the 240 kDa or 230 kDa protein. From these results, it was surmised that there are two species of the rat telomerase protein, one species constituted by a 240 kDa protein component (referred to below as "p240"), and one species constituted by a 230 kDa protein component (referred to below as "p230").

Working Example 9: Production and activation of the rat telomerase molecular species

In order to investigate the generative process for p240 and p230, a cellular pulse-chase experiment was carried out. Cells from the rat hepatoma cell line AH66F held in a 10 cm plastic dish were pulse labeled for 30 min in 1 mL of DMEM medium (lacking methionine and cysteine, manufactured by Life Technologies) containing 250 μCi/mL ^{35}S-methionine (Product name Tran 35S label, ICN) and 10% fetal calf serum (JRH Bioscience), whereupon a large excess of non-radioactive methionine was added to the medium. At 0, 1, 3 and 6 h after addition of the non-radioactive methionine, the cells were collected, and immunoprecipitation was carried out in the same manner as in working Example 4 using antibody specific for recombinant rat telomerase protein.

The immunoprecipitation product obtained in this manner was subjected to 6% SDS PAGE followed by autoradiography. The results indicated that the protein that was immunoprecipitated immediately after pulse labeling (0 h) was primarily p240, but that the amount of p240 decreased and the amount of p230 increased over time (1, 3, 6 h). From this result, it was concluded that the rat telomerase protein is expressed first as a protein with a structure that includes p240, and modification then occurs to produce a protein with a structure that includes p230.

The p240/p230 ratio was also determined in cells from the rat hepatoma cell line AH66F and in normal rat tissue in order to investigate the relationship between this ratio and telomerase activity. First, S100 extracts were prepared following the method of Counter et al. (EMBO J., 11, 1021, 1995) from rat liver, kidney and testes, and from AH66F cells. This extract was partially purified on a heparin sepharose CL-6B column in the same manner as in Working Example 6. For each of the partially purified telomerase fractions, the p230/p240 ratio was determined by western blotting using antibody specific for recombinant rat telomerase protein, and the telomerase activity was measured. The results indicated a decreasing level of telomerase activity in the order: AH66F cells, testes cells, liver cells. No activity was detected in kidney cells. On the other hand, the p230 ratio decreased in the order: AH66F cells, testes cells, liver cells. Almost no p230 was found in kidney cells.

The results indicate a strong relationship between p230 presence and telomerase activity, which leads to the conclusion that p230 is the active form and p240 is the inactive form of the molecular species that constitutes rat telomerase protein. Based on the above, it was confirmed that rat telomerase protein is first produced with a molecular species that is constituted by the inactive-form p240, and modification then occurs to convert p240 into p230, thereby producing the active molecular species.

Possibilities for industrial utilization

In the present invention, a telomerase protein derived from higher animals and a gene encoding the same are offered, which telomerase protein is required for cellular proliferation and has been implicated as having a role in cancer cell proliferation. The telomerase protein and gene encoding the same are expected to be useful in understanding biological control mechanisms such as cellular growth and senescence, and are expected to be particularly useful in the development of cancer therapies. In addition, antibody that specifically recognizes the telomerase protein of the present invention is expected to be useful as a reagent for detecting cancer cells, and as an agent for diagnostic assays aimed at the early detection of cancer. Moreover, because the telomerase protein of the present invention has a subunit with an active and inactive form, this differential molecular weight characteristic can be employed to screen for drugs that act on the telomerase protein using SDS polyacrylamide electrophoresis.

SEQUENCE TABLE

Sequence no.: 1

Sequence length: nucleic acid = 8215, amino acid = 2629

Sequence form: Nucleic acid and amino acid

Topology: Linear double-stranded

Sequence type: cDNA

Origin: Animal name - rat

Sequence:

//insert sequence, pp.50-68//

Sequence no.: 2

Sequence length: nucleic acid = 487, amino acid = 162

Sequence form: Nucleic acid and amino acid

Topology: Linear double-stranded

Sequence type: cDNA

Origin: Animal name - Human

Sequence:

//insert sequence, pp.68-69//

Sequence no.: 3

Sequence length: 347

Sequence form: Nucleic acid

Strand no.: Double-stranded

Topology: Linear

Origin: Animal name - Rat

Sequence:

//insert sequence, pp.69-70//

Sequence no.: 4

Sequence length: 408

Sequence form: Nucleic acid

Strand no.: Double-stranded

Topology: Linear

Origin: Animal name - Rat

Immediate origin: Plasmid RaPC53

Sequence:

//insert sequence, pp.70-71//

Sequence no.: 5

Sequence length: 17

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Other information: R denotes A or G, Y denotes C or T

Sequence:

CARTTYGAYG ARTAYCA

Sequence no.: 6

Sequence length: 17

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Other information: R denotes A or G, N denotes A, G, C or T, W denotes A or T.

Sequence:

ARCATNGCCA TRWANGG

Sequence no.: 7

Sequence length: 23

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Other information: R denotes SA [sic] or G. Y denotes C or T, I denotes inosine

Sequence:

AARTTYGCIC ARTTYGAYGA RTA

Sequence no.: 8

Sequence length: 26

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Other information: R denotes A or G, Y denotes C or T, I denotes inosine

Sequence:

TTYGAYGART AYCARYTIGC IAARTA

Sequence no.: 9

Sequence length: 26

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Other information: R denotes A or G, I denotes inosine, K denotes G or T

Sequence:

ARRTTICKIA RCATIGCCAT RAAIGG

Sequence no.: 10

Sequence length: 26

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Other information: R denotes A or G, I denotes inosine, K denotes G or T

Sequence:

TTRCAIARRT TICKIARCAT IGCCAT

Sequence no.: 11

Sequence length: 23

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Sequence:

CAGGGATGGA GCCTCCATTT TCT

Sequence no.: 12

Sequence length: 23

Sequence form: Nucleic acid

Strand no.: Single-stranded

Topology: Linear

Sequence type: Non-native nucleic acids - Synthetic DNA

Sequence:

TCAATGAGTT CCTCCCAGAC CGA

Sequence no.: 13

Sequence length: nucleic acid = 8839, amino acid = 2625

Sequence form: Nucleic acid and amino acid

Topology: Linear double-stranded

Sequence type: cDNA

Origin: Animal name Human

Sequence:

//insert sequence, pp.74-93//

Claims

1. A polypeptide specified by the amino acid sequence of sequence no. 1 in the sequence table.

2. The polypeptide according to Claim 1, which is a telomerase protein derived from rat.

3. A polypeptide, characterized by having one or more amino acid substitutions, insertions or deletions in the amino acid sequence of sequence no. 1 in the sequence table, which essentially functions as the telomerase protein of higher animals including humans.

4. The polypeptide of Claim 3, that can function as the telomerase protein in the human body.

5. A polypeptide specified by the amino acid sequence of sequence no. 2 in the sequence table.

6. The polypeptide according to Claim 5, which is a partial polypeptide of human telomerase protein.

7. A polypeptide, characterized by having one or more amino acid substitutions, insertions or deletions in the amino acid sequence of sequence no. 2 in the sequence table, which essentially functions as a telomerase protein for higher animals including humans.

8. A polypeptide, specified by the amino acid sequence of sequence no. 13 in the sequence table.

9. The polypeptide according to Claim 8, which is a human telomerase protein.

10. A polypeptide, characterized by having one or more amino acid substitutions, insertions or deletions in the amino acid sequence of sequence no. 13 in the sequence table, which essentially functions as a telomerase protein of higher animals including humans.

11. The polypeptide according to Claim 10, which can function as telomerase protein in the human body.

12. A nucleotide sequence coding for the polypeptide of any of claims 1-11.

13. The nucleotide sequence according to Claim 12, which is a DNA sequence or RNA sequence.

14. A recombinant vector containing the DNA sequence of Claim 13.

15. A transformant containing the recombinant vector of Claim 14.

16. A method for manufacturing the polypeptides of any of Claims 1-11, which includes a process wherein a polypeptide that is the gene product of the DNA sequence of Claim 13 is isolated and collected from a culture produced by culturing the transformant described in Claim 15.

17. An antibody that can specifically recognize the polypeptides of any of Claims 1-11.

18. A nucleic acid probe containing a nucleotide that can bind complementary to all or part of the nucleotide sequence of Claim 12.

19. A reagent for detecting cancer cells, which contains the nucleic acid probe described in Claim 18 or the antibody described in Claim 17.

20. A medical composition for use in cancer diagnosis that contains the nucleic acid probe described in Claim 18 or the antibody described in Claim 17.

21. A higher animal telomerase protein that includes, as a subunit, the polypeptide described in Claim 3 or Claim 10.

22. The polypeptide according to Claim 3, characterized by having a molecular weight, as determined by SDS polyacrylamide electrophoresis, of about 240 kDa in its inactive form, and about 230 kDa in its inactive form.

23. The active-form polypeptide according to Claim 3, characterized by having a molecular weight of about 230 kDa as determined by SDS polyacrylamide electrophoresis.

24. A screening method for substances that act on the expression of enzymatic activity of higher animal telomerase protein, where said screening method includes a process wherein the molecular weight is determined for the telomerase protein, or a subunit thereof, that is contained in cells or tissues that have been in contact with a test substance.

25. The screening method according to Claim 24, wherein the process that involves contact with the test substance is performed by culturing in the presence of the test substance, or administering the test substance to an animal.

26. The screening method according to Claim 24 or 25, wherein molecular weight measurement is carried out by SDS polyacrylamide electrophoresis.

27. The screening method according to Claim 26, that includes a process wherein the ratio of the approximately 240 kDa inactive-form polypeptide and the approximately 230 kDa active-form polypeptide is determined.

28. The screening method according to Claim 26 or 27, which includes a process where the test substance is determined to be a substance that inhibits the expression of enzymatic activity of higher animal telomerase protein when the ratio of the 240 kDa polypeptide is essentially increased in the presence of test substance relative to the ratio of said polypeptide in the absence of the test substance.

29. The screening method according to Claim 26 or 27, which includes a process where the test substance is determined to be a substance that activates the expression of enzymatic activity of higher animal telomerase protein when the ratio of the 230 kDa polypeptide is essentially increased in the presence of test substance relative to the ratio of said polypeptide in the absence of the test substance.

30. The screening method according to any of Claims 24-29, which includes a process for measuring the molecular weight of the polypeptide described in Claim 1 or Claim 3.

Figure 1

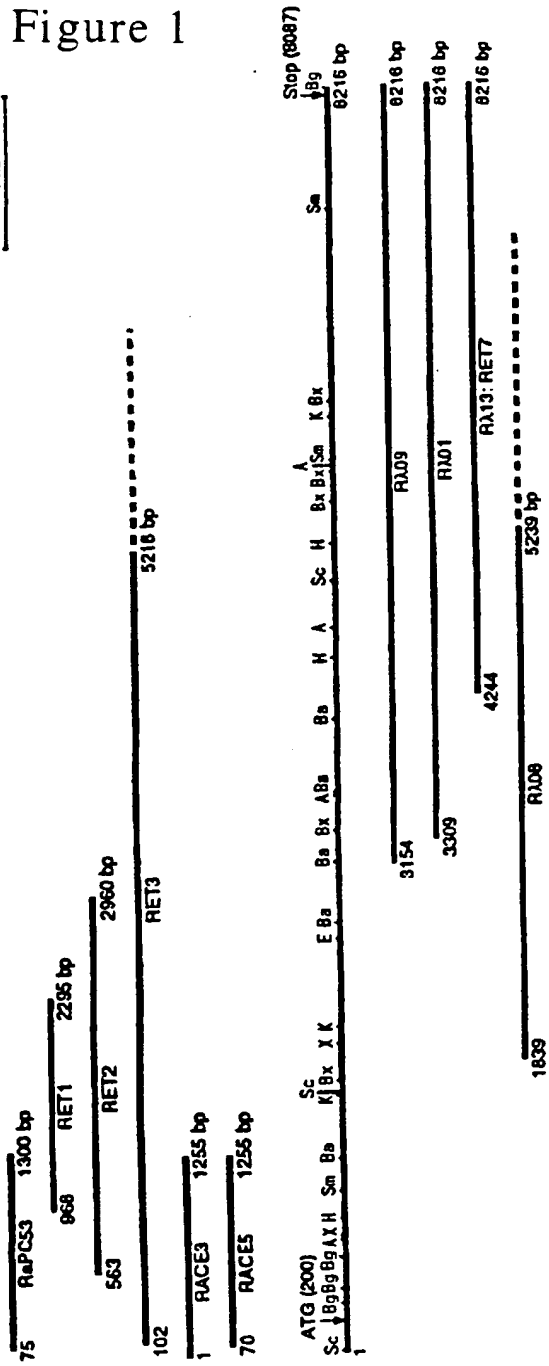

1 kb
RaPC53
1300 bp
RET1
2295 bp
RET2
2960 bp
RET3
5216 bp
RACE3
1255 bp
RACE5
1255 bp
75
868
563
102
1
70
ATG (200)
Stop (8087)
8216 bp
8216 bp
8216 bp
8216 bp
RA09
RA01
RA13: RET7
RA08
5239 bp
3154
3309
4244
1839
1

Figure 2

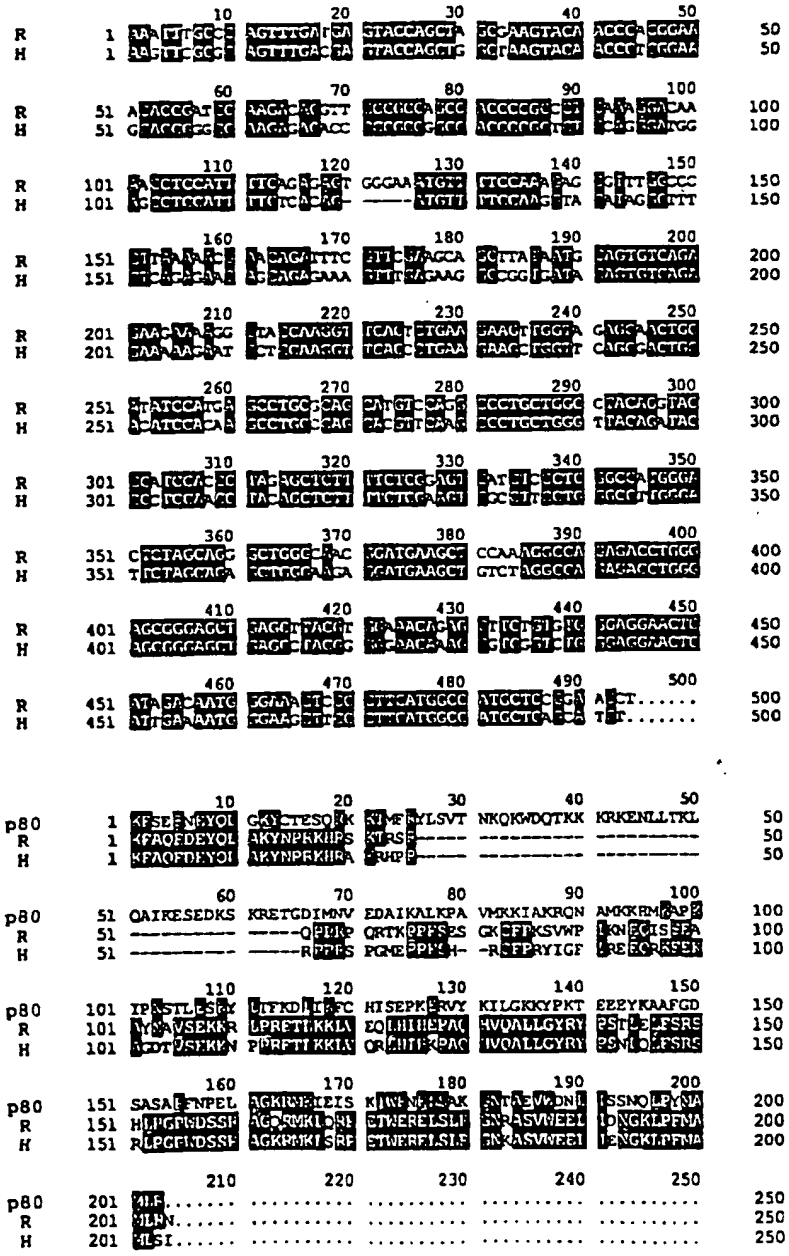


Figure 3

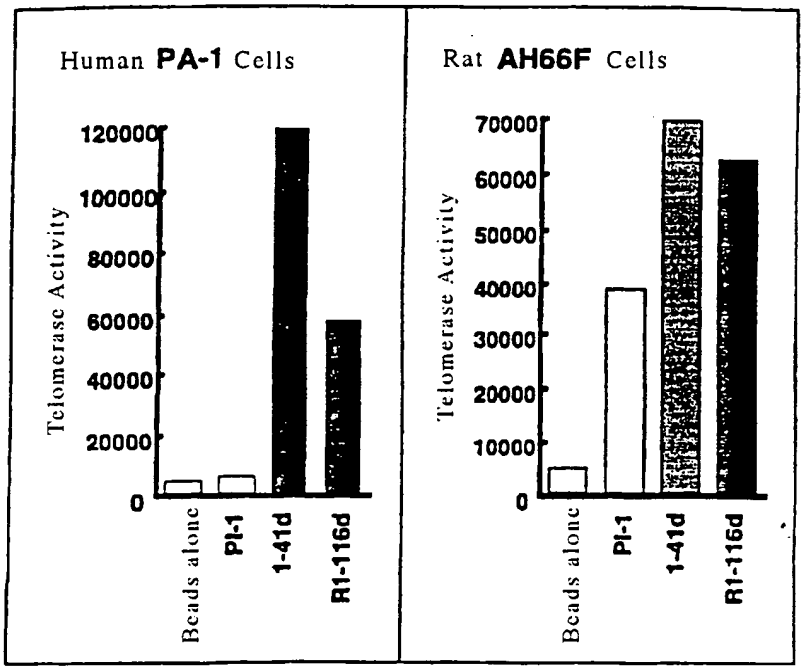
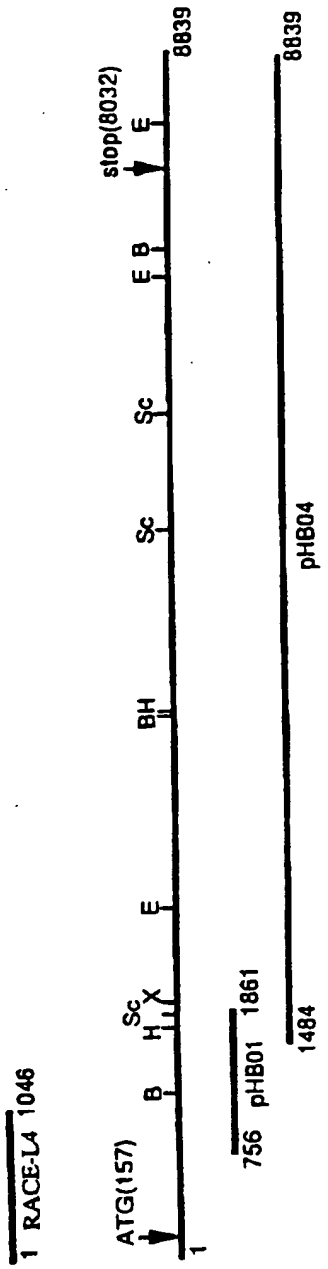

Human PA-1 Cells
Telomerase Activity
120000
100000
80000
60000
40000
20000
0
Beads alone
PI-1
1-41d
R1-116d
Rat AH66F Cells
Telomerase Activity
70000
60000
50000
40000
30000
20000
10000
0
Beads alone
PI-1
1-41d
R1-116d

Figure 4


1 kb
1 RACE-L4 1046
ATG(157)
B
Sc
H
X
E
BH
Sc
Sc
E B
stop(8032)
E
8839
1
756 pHB01 1861
1484
pHB04
8839

配列番号：１３

配列の長さ：核酸＝８８３９、アミノ酸＝２６２５

配列の型：核酸及びアミノ酸

トポロジー：直鎖状二本鎖

配列の種類：ｃＤＮＡ

起源：生物名：ヒト

配列

AGATCCGCAT CCGGCGCCTC CCCCGGCTGC CACCCTTCCC ACCGGCAGAA TCCAGAGCGA 60

AGTTTCTGCT TCCTGCTGCG GGAATCGGAC GCCCCAGGTC AGGCACCCAG GGTTTCCAGC 120

CCCAGTCTAA GGCATATACA AGCTGAGTTT CAGCC ATG GAA AAA CTC CAT 170
Met Glu Lys Leu His
1 5

GGG CAT GTG TCT GCC CAT CCA GAC ATC CTC TCC TTG GAG AAC CGG TGC 218
Gly His Val Ser Ala His Pro Asp Ile Leu Ser Leu Glu Asn Arg Cys
10 15 20

CTG GCT ATG CTC CCT GAC TTA CAG CCC TTG GAG AAA CTA CAT CAG CAT 266
Leu Ala Met Leu Pro Asp Leu Gln Pro Leu Glu Lys Leu His Gln His
25 30 35

GTA TCT ACC CAC TCA GAT ATC CTC TCC TTG AAG AAC CAG TGC CTA GCC 314
Val Ser Thr His Ser Asp Ile Leu Ser Leu Lys Asn Gln Cys Leu Ala
40 45 50

ACG CTT CCT GAC CTG AAG ACC ATG GAA AAA CCA CAT GGA TAT GTG TCT 362
Thr Leu Pro Asp Leu Lys Thr Met Glu Lys Pro His Gly Tyr Val Ser
55 60 65

GCC CAC CCA GAC ATC CTC TCC TTG GAG AAC CAG TGC CTG GCC ACA CTT 410
Ala His Pro Asp Ile Leu Ser Leu Glu Asn Gln Cys Leu Ala Thr Leu
70 75 80 85

TCT GAC CTG AAG ACC ATG GAG AAA CCA CAT GGA CAT GTT TCT GCC CAC 458
Ser Asp Leu Lys Thr Met Glu Lys Pro His Gly His Val Ser Ala His
90 95 100

CCA GAC ATC CTC TCC TTG GAG AAC CGA TGC CTG GCC ACC CTC TCT AGT 506
Pro Asp Ile Leu Ser Leu Glu Asn Arg Cys Leu Ala Thr Leu Ser Ser
105 110 115

CTA AAG AGC ACT GTG TCT GCC AGC CCC TTG TTC CAG AGT CTA CAG ATA 554
Leu Lys Ser Thr Val Ser Ala Ser Pro Leu Phe Gln Ser Leu Gln Ile
120 125 130

TCT CAC ATG ATG CAA GCT GAT TTG TAC CGT GTG AAC AAC AGC AAT TGC 602

```
Ser His Met Met Gln Ala Asp Leu Tyr Arg Val Asn Asn Ser Asn Cys
    135                 140                 145
CTG CTC TCT GAG CCT CCA AGT TGG AGG GCT CAG CAT TTC TCT AAG GGA    650
Leu Leu Ser Glu Pro Pro Ser Trp Arg Ala Gln His Phe Ser Lys Gly
150                 155                 160                 165
CTA GAC CTT TCA ACC TGC CCT ATA GCC CTG AAA TCC ATC TCT GCC ACA    698
Leu Asp Leu Ser Thr Cys Pro Ile Ala Leu Lys Ser Ile Ser Ala Thr
                170                 175                 180
GAG ACA GCT CAG GAA GCA ACT TTG GGT CGT TGG TTT GAT TCA GAA GAG    746
Glu Thr Ala Gln Glu Ala Thr Leu Gly Arg Trp Phe Asp Ser Glu Glu
            185                 190                 195
AAG AAA GGG GCA GAG ACC CAA ATG CCT TCT TAT AGT CTG AGC TTG GGA    794
Lys Lys Gly Ala Glu Thr Gln Met Pro Ser Tyr Ser Leu Ser Leu Gly
        200                 205                 210
GAG GAG GAG GAG GTG GAG GAT CTG GCC GTG AAG CTC ACC TCT GGA GAC    842
Glu Glu Glu Glu Val Glu Asp Leu Ala Val Lys Leu Thr Ser Gly Asp
    215                 220                 225
TCT GAA TCT CAT CCA GAG CCT ACT GAC CAT GTC CTT CAG GAA AAG AAG    890
Ser Glu Ser His Pro Glu Pro Thr Asp His Val Leu Gln Glu Lys Lys
230                 235                 240                 245
ATG GCT CTA CTG AGC TTG CTG TGC TCT ACT CTG GTC TCA GAA GTA AAC    938
Met Ala Leu Leu Ser Leu Leu Cys Ser Thr Leu Val Ser Glu Val Asn
                250                 255                 260
ATG AAC AAT ACA TCT GAC CCC ACC CTG GCT GCC ATT TTT GAA ATC TGT    986
Met Asn Asn Thr Ser Asp Pro Thr Leu Ala Ala Ile Phe Glu Ile Cys
            265                 270                 275
CGT GAA CTT GCC CTC CTG GAG CCT GAG TTT ATC CTC AAG GCA TCT TTG   1034
Arg Glu Leu Ala Leu Leu Glu Pro Glu Phe Ile Leu Lys Ala Ser Leu
```

```
        280                 285                 290
TAT GCC AGG CAG CAG CTG AAC GTC CGG AAT GTG GCC AAT AAA ATC TTG   1082
Tyr Ala Arg Gln Gln Leu Asn Val Arg Asn Val Ala Asn Lys Ile Leu
    295                 300                 305
GCC ATT GCT GCT TTC TTG CCG GCG TGT CGC CCC CAC CTG CGA CGA TAT   1130
Ala Ile Ala Ala Phe Leu Pro Ala Cys Arg Pro His Leu Arg Arg Tyr
310                 315                 320                 325
TTC TGT GCC ATT GTC CAG CTG CCT TCT GAC TGG ATC CAG GTG GCT GAG   1178
Phe Cys Ala Ile Val Gln Leu Pro Ser Asp Trp Ile Gln Val Ala Glu
                330                 335                 340
CTT TAC CAG AGC CTG GCT GAG GGA GAT AAG AAT AAG CTG GTG CCC CTG   1226
Leu Tyr Gln Ser Leu Ala Glu Gly Asp Lys Asn Lys Leu Val Pro Leu
            345                 350                 355
CCC GCC TGT CTC CGT ACT GCC ATG ACG GAC AAA TTT GCC CAG TTT GAC   1274
Pro Ala Cys Leu Arg Thr Ala Met Thr Asp Lys Phe Ala Gln Phe Asp
        360                 365                 370
GAG TAC CAG CTG GCT AAG TAC AAC CCT CGG AAG CAC CGG GCC AAG AGA   1322
Glu Tyr Gln Leu Ala Lys Tyr Asn Pro Arg Lys His Arg Ala Lys Arg
    375                 380                 385
CAC CCC CGC CGG CCA CCC CGC TCT CCA GGG ATG GAG CCT CCA TTT TCT   1370
His Pro Arg Arg Pro Pro Arg Ser Pro Gly Met Glu Pro Pro Phe Ser
390                 395                 400                 405
CAC AGA TGT TTT CCA AGG TAC ATA GGG TTT CTC AGA GAA GAG CAG AGA   1418
His Arg Cys Phe Pro Arg Tyr Ile Gly Phe Leu Arg Glu Glu Gln Arg
                410                 415                 420
AAG TTT GAG AAG GCC GGT GAT ACA GTG TCA GAG AAA AAG AAT CCT CCA   1466
Lys Phe Glu Lys Ala Gly Asp Thr Val Ser Glu Lys Lys Asn Pro Pro
            425                 430                 435
```

AGG TTC ACC CTG AAG AAG CTG GTT CAG CGA CTG CAC ATC CAC AAG CCT 1514
Arg Phe Thr Leu Lys Lys Leu Val Gln Arg Leu His Ile His Lys Pro
440 445 450

GCC CAG CAC GTT CAA GCC CTG CTG GGT TAC AGA TAC CCC TCC AAC CTA 1562
Ala Gln His Val Gln Ala Leu Leu Gly Tyr Arg Tyr Pro Ser Asn Leu
455 460 465

CAG CTC TTT TCT CGA AGT CGC CTT CCT GGG CCT TGG GAT TCT AGC AGA 1610
Gln Leu Phe Ser Arg Ser Arg Leu Pro Gly Pro Trp Asp Ser Ser Arg
470 475 480 485

GCT GGG AAG AGG ATG AAG CTG TCT AGG CCA GAG ACC TGG GAG CGG GAG 1658
Ala Gly Lys Arg Met Lys Leu Ser Arg Pro Glu Thr Trp Glu Arg Glu
490 495 500

CTG AGC CTA CGG GGG AAC AAA GCG TCG GTC TGG GAG GAA CTC ATT GAA 1706
Leu Ser Leu Arg Gly Asn Lys Ala Ser Val Trp Glu Glu Leu Ile Glu
505 510 515

AAT GGG AAG CTT CCC TTC ATG GCC ATG CTT CGG AAC CTG TGC AAC CTG 1754
Asn Gly Lys Leu Pro Phe Met Ala Met Leu Arg Asn Leu Cys Asn Leu
520 525 530

CTG CGG GTT GGA ATC AGT TCC CGC CAC CAT GAG CTC ATT CTC CAG AGA 1802
Leu Arg Val Gly Ile Ser Ser Arg His His Glu Leu Ile Leu Gln Arg
535 540 545

CTC CAG CAT GCG AAG TCG GTG ATC CAC AGT CGG CAG TTT CCA TTC AGA 1850
Leu Gln His Ala Lys Ser Val Ile His Ser Arg Gln Phe Pro Phe Arg
550 555 560 565

TTT CTT AAC GCC CAT GAT GCC ATT GAT GCC CTC GAG GCT CAA CTC AGA 1898
Phe Leu Asn Ala His Asp Ala Ile Asp Ala Leu Glu Ala Gln Leu Arg
570 575 580

AAT CAA GCA TTG CCC TTT CCT TCG AAT ATA ACA CTG ATG AGG CGG ATA 1946

```
Asn Gln Ala Leu Pro Phe Pro Ser Asn Ile Thr Leu Met Arg Arg Ile
            585                 590                 595
CTA ACT AGA AAT GAA AAG AAC CGT CCC AGG CGG AGG TTT CTT TGC CAC   1994
Leu Thr Arg Asn Glu Lys Asn Arg Pro Arg Arg Arg Phe Leu Cys His
        600                 605                 610
CTA AGC CGT CAG CAG CTT CGG ATG GCA ATG AGG ATA CCT GTG TTG TAT   2042
Leu Ser Arg Gln Gln Leu Arg Met Ala Met Arg Ile Pro Val Leu Tyr
    615                 620                 625
GAG CAG CTC AAG AGG GAG AAG CTG AGA GTA CAC AAG GCC AGA CAG TGG   2090
Glu Gln Leu Lys Arg Glu Lys Leu Arg Val His Lys Ala Arg Gln Trp
630                 635                 640                 645
AAA TAT GAT GGT GAG ATG CTG AAC AGG TAC CGA CAG GCC CTA GAG ACA   2138
Lys Tyr Asp Gly Glu Met Leu Asn Arg Tyr Arg Gln Ala Leu Glu Thr
                650                 655                 660
GCT GTG AAC CTC TCT GTG AAG CAC AGC CTG CCC CTG CTG CCA GGC CGC   2186
Ala Val Asn Leu Ser Val Lys His Ser Leu Pro Leu Leu Pro Gly Arg
            665                 670                 675
ACT GTC TTG GTC TAT CTG ACA GAT GCT AAT GCA GAC AGG CTC TGT CCA   2234
Thr Val Leu Val Tyr Leu Thr Asp Ala Asn Ala Asp Arg Leu Cys Pro
        680                 685                 690
AAG AGC AAC CCA CAA GGG CCC CCG CTG AAC TAT GCA CTG CTG TTG ATT   2282
Lys Ser Asn Pro Gln Gly Pro Pro Leu Asn Tyr Ala Leu Leu Leu Ile
    695                 700                 705
GGG ATG ATG ATC ACG AGG GCG GAG CAG GTG GAC GTC GTG CTG TGT GGA   2330
Gly Met Met Ile Thr Arg Ala Glu Gln Val Asp Val Val Leu Cys Gly
710                 715                 720                 725
GGT GAC ACT CTG AAG ACT GCA GTG CTT AAG GCA GAA GAA GGC ATC CTG   2378
Gly Asp Thr Leu Lys Thr Ala Val Leu Lys Ala Glu Glu Gly Ile Leu
```

730 735 740

AAG ACT GCC ATC AAG CTC CAG GCT CAA GTC CAG GAG TTT GAT GAA AAT 2426
Lys Thr Ala Ile Lys Leu Gln Ala Gln Val Gln Glu Phe Asp Glu Asn

745 750 755

GAT GGA TGG TCC CTG AAT ACT TTT GGG AAA TAC CTG CTG TCT CTG GCT 2474
Asp Gly Trp Ser Leu Asn Thr Phe Gly Lys Tyr Leu Leu Ser Leu Ala

760 765 770

GGC CAA AGG GTT CCT GTG GAC AGG GTC ATC CTC CTT GGC CAA AGC ATG 2522
Gly Gln Arg Val Pro Val Asp Arg Val Ile Leu Leu Gly Gln Ser Met

775 780 785

GAT GAT GGA ATG ATA AAT GTG GCC AAA CAG CTT TAC TGG CAG CGT GTG 2570
Asp Asp Gly Met Ile Asn Val Ala Lys Gln Leu Tyr Trp Gln Arg Val

790 795 800 805

AAT TCC AAG TGC CTC TTT GTT GGT ATC CTC CTA AGA AGG GTA CAA TAC 2618
Asn Ser Lys Cys Leu Phe Val Gly Ile Leu Leu Arg Arg Val Gln Tyr

810 815 820

CTG TCA ACA GAT TTG AAT CCC AAT GAT GTG ACA CTC TCA GGC TGT ACT 2666
Leu Ser Thr Asp Leu Asn Pro Asn Asp Val Thr Leu Ser Gly Cys Thr

825 830 835

GAT GCG ATA CTG AAG TTC ATT GCA GAG CAT GGG GCC TCC CAT CTT CTG 2714
Asp Ala Ile Leu Lys Phe Ile Ala Glu His Gly Ala Ser His Leu Leu

840 845 850

GAA CAT GTG GGC CAA ATG GAC AAA ATA TTC AAG ATT CCA CCA CCC CCA 2762
Glu His Val Gly Gln Met Asp Lys Ile Phe Lys Ile Pro Pro Pro Pro

855 860 865

GGA AAG ACA GGG GTC CAG TCT CTC CGG CCA CTG GAA GAG GAC ACT CCA 2810
Gly Lys Thr Gly Val Gln Ser Leu Arg Pro Leu Glu Glu Asp Thr Pro

870 875 880 885

```
AGC CCC TTG GCT CCT GTT TCC CAG CAA GGA TGG GGC AGC ATC CGG CTT   2858
Ser Pro Leu Ala Pro Val Ser Gln Gln Gly Trp Gly Ser Ile Arg Leu
                890                 895                 900
TTC ATT TCA TCC ACT TTC CGA GAC ATG CAC CGG GGA GCG GAC CTG CTG   2906
Phe Ile Ser Ser Thr Phe Arg Asp Met His Arg Gly Ala Asp Leu Leu
            905                 910                 915
CTG AGG TCT GTG CTG CCA GCA CTG CAG GCC CGA GCG GCC CCT CAC CGT   2954
Leu Arg Ser Val Leu Pro Ala Leu Gln Ala Arg Ala Ala Pro His Arg
        920                 925                 930
ATC AGC CTT CAC CGA ATC GAC CTC CGC TGG GGC GTC ACT GAG GAG GAG   3002
Ile Ser Leu His Arg Ile Asp Leu Arg Trp Gly Val Thr Glu Glu Glu
    935                 940                 945
ACC CGT AGG AAC AGA CAA CTG GAA GTG TGC CTT GGG GAG GTG GAG AAC   3050
Thr Arg Arg Asn Arg Gln Leu Glu Val Cys Leu Gly Glu Val Glu Asn
950                 955                 960                 965
GCA CAG CTG TTT GTG GGG ATT CTG GGC TCC CGT TAT GGA AAC ATT CCC   3098
Ala Gln Leu Phe Val Gly Ile Leu Gly Ser Arg Tyr Gly Asn Ile Pro
                970                 975                 980
CCC AGC TAC AAC CTT CCT GAC CAT CCA CAC TTC CAC TGG GCC CAG CAG   3146
Pro Ser Tyr Asn Leu Pro Asp His Pro His Phe His Trp Ala Gln Gln
            985                 990                 995
TAC CCT TCA GGG CGC TCT GTG ACA GAG ATG GAG GTG ATG CAG TTC CTG   3194
Tyr Pro Ser Gly Arg Ser Val Thr Glu Met Glu Val Met Gln Phe Leu
        1000                1005                1010
AAC CGG AAC CAA CGT CTG CAG CCC TCT GCC CAA GCT CTC ATC TAC TTC   3242
Asn Arg Asn Gln Arg Leu Gln Pro Ser Ala Gln Ala Leu Ile Tyr Phe
    1015                1020                1025
CGG GAT TCC AGC TTC CTC AGC TCT GTG CCA GAT GCC TGG AAA TCT GAC   3290
```

Arg Asp Ser Ser Phe Leu Ser Ser Val Pro Asp Ala Trp Lys Ser Asp

1030 1035 1040 1045

TTT GTT TCT GAG TCT GAA GAG GCC GCA TGT CGG ATC TCA GAA CTG AAG 3338

Phe Val Ser Glu Ser Glu Glu Ala Ala Cys Arg Ile Ser Glu Leu Lys

1050 1055 1060

AGC TAC CTA AGC AGA CAG AAA GGG ATA ACC TGC CGC AGA TAC CCC TGT 3386

Ser Tyr Leu Ser Arg Gln Lys Gly Ile Thr Cys Arg Arg Tyr Pro Cys

1065 1070 1075

GAG TGG GGG GGT GTG GCA GCT GGC CGG CCC TAT GTT GGC GGG CTG GAG 3434

Glu Trp Gly Gly Val Ala Ala Gly Arg Pro Tyr Val Gly Gly Leu Glu

1080 1085 1090

GAG TTT GGG CAG TTG GTT CTG CAG GAT GTA TGG AAT ATG ATC CAG AAG 3482

Glu Phe Gly Gln Leu Val Leu Gln Asp Val Trp Asn Met Ile Gln Lys

1095 1100 1105

CTC TAC CTG CAG CCT GGG GCC CTG CTG GAG CAG CCA GTG TCC ATC CCA 3530

Leu Tyr Leu Gln Pro Gly Ala Leu Leu Glu Gln Pro Val Ser Ile Pro

1110 1115 1120 1125

GAC GAT GAC TTG GTC CAG GCC ACC TTC CAG CAG CTG CAG AAG CCA CCG 3578

Asp Asp Asp Leu Val Gln Ala Thr Phe Gln Gln Leu Gln Lys Pro Pro

1130 1135 1140

AGT CCT GCC CGG CCA CGC CTT CTT CAG GAC ACA GTG CAA CGG CTG ATG 3626

Ser Pro Ala Arg Pro Arg Leu Leu Gln Asp Thr Val Gln Arg Leu Met

1145 1150 1155

CTG CCC CAC GGA AGG CTG AGC CTG GTG ACG GGG CAG TCA GGA CAG GGC 3674

Leu Pro His Gly Arg Leu Ser Leu Val Thr Gly Gln Ser Gly Gln Gly

1160 1165 1170

AAG ACA GCC TTC CTG GCA TCT CTT GTG TCA GCC CTG CAG GCT CCT GAT 3722

Lys Thr Ala Phe Leu Ala Ser Leu Val Ser Ala Leu Gln Ala Pro Asp

1175 1180 1185

GGG GCC AAG GTG GCA CCA TTA GTC TTC TTC CAC TTT TCT GGG GCT CGT 3770

Gly Ala Lys Val Ala Pro Leu Val Phe Phe His Phe Ser Gly Ala Arg

1190 1195 1200 1205

CCT GAC CAG GGT CTT GCC CTC ACT CTG CTC AGA CGC CTC TGT ACC TAT 3818

Pro Asp Gln Gly Leu Ala Leu Thr Leu Leu Arg Arg Leu Cys Thr Tyr

1210 1215 1220

CTG CGT GGC CAA CTA AAA GAG TCA GGT GCC CTC CCC AGC ACC TAC CGA 3866

Leu Arg Gly Gln Leu Lys Glu Ser Gly Ala Leu Pro Ser Thr Tyr Arg

1225 1230 1235

AGC CTG GTG TGG GAG CTG CAG CAG AGG CTG CTG CCC AAG TCT GCT GAG 3914

Ser Leu Val Trp Glu Leu Gln Gln Arg Leu Leu Pro Lys Ser Ala Glu

1240 1245 1250

TCC CTG CAT CCT GGC CAG ACC CAG GTC CTG ATC ATC GAT GGG GCT GAT 3962

Ser Leu His Pro Gly Gln Thr Gln Val Leu Ile Ile Asp Gly Ala Asp

1255 1260 1265

AGG TTA GTG GAC CAG AAT GGG CAG CTG ATT TCA GAC TGG ATC CCA AAG 4010

Arg Leu Val Asp Gln Asn Gly Gln Leu Ile Ser Asp Trp Ile Pro Lys

1270 1275 1280 1285

AAG CTT CCC CGG TGT GTA CAC CTG GTG CTG AGT GTG TCT AGT GAT GCA 4058

Lys Leu Pro Arg Cys Val His Leu Val Leu Ser Val Ser Ser Asp Ala

1290 1295 1300

GGC CTA GGG GAG ACC CTT GAG CAG AGC CAG GGT GCC CAC GTG CTG GCC 4106

Gly Leu Gly Glu Thr Leu Glu Gln Ser Gln Gly Ala His Val Leu Ala

1305 1310 1315

TTG GGG CCT CTG GAG GCC TCT GCT CGG GCC CGG CTG GTG AGA GAG GAG 4154

Leu Gly Pro Leu Glu Ala Ser Ala Arg Ala Arg Leu Val Arg Glu Glu

1320 1325 1330

CTG GCC CTG TAC GGG AAG CGG CTG GAG GAG TCA CCA TTT AAC AAC CAG 4202
Leu Ala Leu Tyr Gly Lys Arg Leu Glu Glu Ser Pro Phe Asn Asn Gln
1335 1340 1345

ATG CGA CTG CTG CTG GTG AAG CGG GAA TCA GGC CGG CCG CTC TAC CTG 4250
Met Arg Leu Leu Leu Val Lys Arg Glu Ser Gly Arg Pro Leu Tyr Leu
1350 1355 1360 1365

CGC TTG GTC ACC GAT CAC CTG AGG CTC TTC ACG CTG TAT GAG CAG GTG 4298
Arg Leu Val Thr Asp His Leu Arg Leu Phe Thr Leu Tyr Glu Gln Val
1370 1375 1380

TCT GAG AGA CTC CGG ACC CTG CCT GCC ACT GTC CCC CTG CTG CAG CAC 4346
Ser Glu Arg Leu Arg Thr Leu Pro Ala Thr Val Pro Leu Leu Gln His
1385 1390 1395

ATC CTG AGC ACA CTG GAG AAG GAG CAC GGG CCT GAT GTC CTT CCC CAG 4394
Ile Leu Ser Thr Leu Glu Lys Glu His Gly Pro Asp Val Leu Pro Gln
1400 1405 1410

GCC TTG ACT GCC CTA GAA GTC ACA CGG AGT GGT TTG ACT GTG GAC CAG 4442
Ala Leu Thr Ala Leu Glu Val Thr Arg Ser Gly Leu Thr Val Asp Gln
1415 1420 1425

CTG CAC GGA GTG CTG AGT GTG TGG CGG ACA CTA CCG AAG GGG ACT AAG 4490
Leu His Gly Val Leu Ser Val Trp Arg Thr Leu Pro Lys Gly Thr Lys
1430 1435 1440 1445

ACC TGG GAA GAA GCA GTG GCT GCT GGT AAC AGT GGA GAC CCC TAC CCC 4538
Thr Trp Glu Glu Ala Val Ala Ala Gly Asn Ser Gly Asp Pro Tyr Pro
1450 1455 1460

ATG GGC CCG TTT GCC TAC CTC GTC CAG AGT CTG CGC AGT TTG CTA GGG 4586
Met Gly Pro Phe Ala Tyr Leu Val Gln Ser Leu Arg Ser Leu Leu Gly
1465 1470 1475

GAG GGC CCT CTG GAG CGC CCT GGT GCC CGG CTG TGC CTC CCT GAT GGG 4634

Glu Gly Pro Leu Glu Arg Pro Gly Ala Arg Leu Cys Leu Pro Asp Gly

1480 1485 1490

CCC CTG AGA ACA GCA GCT AAA CGT TGC TAT GGG AAG AGG CCA GGG CTA 4682

Pro Leu Arg Thr Ala Ala Lys Arg Cys Tyr Gly Lys Arg Pro Gly Leu

1495 1500 1505

GAG GAC ACG GCA CAC ATC CTC ATT GCA GCT CAG CTC TGG AAG ACA TGT 4730

Glu Asp Thr Ala His Ile Leu Ile Ala Ala Gln Leu Trp Lys Thr Cys

1510 1515 1520 1525

GAC GCT GAT GCC TCA GGC ACC TTC CGA AGT TGC CCT CCT GAG GCT CTG 4778

Asp Ala Asp Ala Ser Gly Thr Phe Arg Ser Cys Pro Pro Glu Ala Leu

1530 1535 1540

GGA GAC CTG CCT TAC CAC CTG CTC CAG AGC GGG AAC CGT GGA CTT CTT 4826

Gly Asp Leu Pro Tyr His Leu Leu Gln Ser Gly Asn Arg Gly Leu Leu

1545 1550 1555

TCG AAG TTC CTT ACC AAC CTC CAT GTG GTG GCT GCA CAC TTG GAA TTG 4874

Ser Lys Phe Leu Thr Asn Leu His Val Val Ala Ala His Leu Glu Leu

1560 1565 1570

GGT CTG GTC TCT CGG CTC TTG GAG GCC CAT GCC CTC TAT GCT TCT TCA 4922

Gly Leu Val Ser Arg Leu Leu Glu Ala His Ala Leu Tyr Ala Ser Ser

1575 1580 1585

GTC CCC AAA GAG GAA CAA AAG CTC CCC GAG GCT GAC GTT GCA GTG TTT 4970

Val Pro Lys Glu Glu Gln Lys Leu Pro Glu Ala Asp Val Ala Val Phe

1590 1595 1600 1605

CGC ACC TTC CTG AGG CAG CAG GCT TCA ATC CTC AGC CAG TAC CCC CGG 5018

Arg Thr Phe Leu Arg Gln Gln Ala Ser Ile Leu Ser Gln Tyr Pro Arg

1610 1615 1620

CTC CTG CCC CAG CAG GCA GCC AAC CAG CCC CTG GAC TCA CCT CTT TGC 5066

Leu Leu Pro Gln Gln Ala Ala Asn Gln Pro Leu Asp Ser Pro Leu Cys

1625 1630 1635

CAC CAA GCC TCG CTG CTC TCC CGG AGA TGG CAC CTC CAA CAC ACA CTA 5114

His Gln Ala Ser Leu Leu Ser Arg Arg Trp His Leu Gln His Thr Leu

1640 1645 1650

CGA TGG CTT AAT AAA CCC CGG ACC ATG AAA AAT CAG CAA AGC TCC AGC 5162

Arg Trp Leu Asn Lys Pro Arg Thr Met Lys Asn Gln Gln Ser Ser Ser

1655 1660 1665

CTG TCT CTG GCA GTT TCC TCA TCC CCT ACT GCT GTG GCC TTC TCC ACC 5210

Leu Ser Leu Ala Val Ser Ser Ser Pro Thr Ala Val Ala Phe Ser Thr

1670 1675 1680 1685

AAT GGG CAA AGA GCA GCT GTG GGC ACT GCC AAT GGG ACA GTT TAC CTG 5258

Asn Gly Gln Arg Ala Ala Val Gly Thr Ala Asn Gly Thr Val Tyr Leu

1690 1695 1700

TTG GAC CTG AGA ACT TGG CAG GAG GAG AAG TCT GTG GTG AGT GGC TGT 5306

Leu Asp Leu Arg Thr Trp Gln Glu Glu Lys Ser Val Val Ser Gly Cys

1705 1710 1715

GAT GGA ATC TCT GCT TGT TTG TTC CTC TCC GAT GAC ACA CTC TTT CTT 5354

Asp Gly Ile Ser Ala Cys Leu Phe Leu Ser Asp Asp Thr Leu Phe Leu

1720 1725 1730

ACT GCC TTC GAC GGG CTC CTG GAG CTC TGG GAC CTG CAG CAT GGT TGT 5402

Thr Ala Phe Asp Gly Leu Leu Glu Leu Trp Asp Leu Gln His Gly Cys

1735 1740 1745

CGG GTG CTG CAG ACT AAG GCT CAC CAG TAC CAA ATC ACT GGC TGC TGC 5450

Arg Val Leu Gln Thr Lys Ala His Gln Tyr Gln Ile Thr Gly Cys Cys

1750 1755 1760 1765

CTG AGC CCA GAC TGC CGG CTG CTA GCC ACC GTG TGC TTG GGA GGA TGC 5498

Leu Ser Pro Asp Cys Arg Leu Leu Ala Thr Val Cys Leu Gly Gly Cys

1770 1775 1780

```
CTA AAG CTG TGG GAC ACA GTC CGT GGG CAG CTG GCC TTC CAG CAC ACC   5546
Leu Lys Leu Trp Asp Thr Val Arg Gly Gln Leu Ala Phe Gln His Thr
          1785                1790                1795
TAC CCC AAG TCC CTG AAC TGT GTT GCC TTC CAC CCA GAG GGG CAG GTA   5594
Tyr Pro Lys Ser Leu Asn Cys Val Ala Phe His Pro Glu Gly Gln Val
      1800                1805                1810
ATA GCC ACA GGC AGC TGG GCT GGC AGC ATC AGC TTC TTC CAG GTG GAT   5642
Ile Ala Thr Gly Ser Trp Ala Gly Ser Ile Ser Phe Phe Gln Val Asp
  1815                1820                1825
GGG CTC AAA GTC ACC AAG GGA CCT GGG GGC CCC GGA GCC TCT ATC CGT   5690
Gly Leu Lys Val Thr Lys Gly Pro Gly Gly Pro Gly Ala Ser Ile Arg
1830                1835                1840                1845
ACC TTG GCC TTC AAT GTG CCT GGG GGG GTT GTG GCT GTG GGC CGG CTG   5738
Thr Leu Ala Phe Asn Val Pro Gly Gly Val Val Ala Val Gly Arg Leu
              1850                1855                1860
GAC AGT ATG GTG GAG CTG TGG GCC TGG CGA GAA GGG GCA CGG CTG GCT   5786
Asp Ser Met Val Glu Leu Trp Ala Trp Arg Glu Gly Ala Arg Leu Ala
          1865                1870                1875
GCC TTC CCT GCC CAC CAT GGC TTT GTT GCT GCT GCG CTT TTC CTG CAT   5834
Ala Phe Pro Ala His His Gly Phe Val Ala Ala Ala Leu Phe Leu His
      1880                1885                1890
GCG GGT TGC CAG TTA CTG ACG GCT GGA GAG GAT GGC AAG GTT CAG GTG   5882
Ala Gly Cys Gln Leu Leu Thr Ala Gly Glu Asp Gly Lys Val Gln Val
  1895                1900                1905
TGG TCA GGG TCT CTG GGT CGG CCC CGT GGG CAC CTG GGT TCC CTT TCT   5930
Trp Ser Gly Ser Leu Gly Arg Pro Arg Gly His Leu Gly Ser Leu Ser
1910                1915                1920                1925
CTC TCT CCT GCC CTC TCT GTG GCA CTC AGC CCA GAT GGT GAT CGG GTG   5978
```

Leu Ser Pro Ala Leu Ser Val Ala Leu Ser Pro Asp Gly Asp Arg Val

1930 1935 1940

GCT GTT GGA TAT CGA GCG GAT GGC ATT AGG ATC TAC AAA ATC TCT TCA 6026

Ala Val Gly Tyr Arg Ala Asp Gly Ile Arg Ile Tyr Lys Ile Ser Ser

1945 1950 1955

GGT TCC CAG GGG GCT CAG GGT CAG GCA CTG GAT GTG GCA GTG TCG GCC 6074

Gly Ser Gln Gly Ala Gln Gly Gln Ala Leu Asp Val Ala Val Ser Ala

1960 1965 1970

CTG GCC TGG ATA AGC CCC AAG GTA TTG GTG AGT GGT GCA GAA GAT GGG 6122

Leu Ala Trp Ile Ser Pro Lys Val Leu Val Ser Gly Ala Glu Asp Gly

1975 1980 1985

TCC TTG CAG GGC TGG GCA CTC AAG GAA TGC TCC CTT CAG TCC CTC TGG 6170

Ser Leu Gln Gly Trp Ala Leu Lys Glu Cys Ser Leu Gln Ser Leu Trp

1990 1995 2000 2005

CTC CTG TCC AGA TTC CAG AAG CCT GTG CTA GGA CTG GCC ACT TCC CAG 6218

Leu Leu Ser Arg Phe Gln Lys Pro Val Leu Gly Leu Ala Thr Ser Gln

2010 2015 2020

GAG CTC TTG GCT TCT GCC TCA GAG GAT TTC ACA GTG CAG CTG TGG CCA 6266

Glu Leu Leu Ala Ser Ala Ser Glu Asp Phe Thr Val Gln Leu Trp Pro

2025 2030 2035

AGG CAG CTG CTG ACG CGG CCA CAC AAG GCA GAA GAC TTT CCC TGT GGC 6314

Arg Gln Leu Leu Thr Arg Pro His Lys Ala Glu Asp Phe Pro Cys Gly

2040 2045 2050

ACT GAG CTG CGG GGA CAT GAG GGC CCT GTG AGC TGC TGT AGT TTC AGC 6362

Thr Glu Leu Arg Gly His Glu Gly Pro Val Ser Cys Cys Ser Phe Ser

2055 2060 2065

ACT GAT GGA GGC AGC CTG GCC ACC GGG GGC CGG GAT CGG AGT CTC CTC 6410

Thr Asp Gly Gly Ser Leu Ala Thr Gly Gly Arg Asp Arg Ser Leu Leu

2070 2075 2080 2085

TGC TGG GAC GTG AGG ACA CCC AAA ACC CCT GTT TTG ATC CAC TCC TTC 6458

Cys Trp Asp Val Arg Thr Pro Lys Thr Pro Val Leu Ile His Ser Phe

2090 2095 2100

CCT GCC TGT CAC CGT GAC TGG GTC ACT GGC TGT GCC TGG ACC AAA GAT 6506

Pro Ala Cys His Arg Asp Trp Val Thr Gly Cys Ala Trp Thr Lys Asp

2105 2110 2115

AAC CTA CTG ATA TCC TGC TCC AGT GAT GGC TCT GTG GGG CTC TGG GAC 6554

Asn Leu Leu Ile Ser Cys Ser Ser Asp Gly Ser Val Gly Leu Trp Asp

2120 2125 2130

CCA GAG TCA GGA CAG CGG CTT GGT CAG TTC CTG GGT CAT CAG AGT GCT 6602

Pro Glu Ser Gly Gln Arg Leu Gly Gln Phe Leu Gly His Gln Ser Ala

2135 2140 2145

GTG AGC GCT GTG GCA GCT GTG GAG GAG CAC GTG GTG TCT GTG AGC CGG 6650

Val Ser Ala Val Ala Ala Val Glu Glu His Val Val Ser Val Ser Arg

2150 2155 2160 2165

GAT GGG ACC TTG AAA GTG TGG GAC CAT CAA GGC GTG GAG CTG ACC AGC 6698

Asp Gly Thr Leu Lys Val Trp Asp His Gln Gly Val Glu Leu Thr Ser

2170 2175 2180

ATC CCT GCT CAC TCA GGA CCC ATT AGC CAC TGT GCA GCT GCC ATG GAG 6746

Ile Pro Ala His Ser Gly Pro Ile Ser His Cys Ala Ala Ala Met Glu

2185 2190 2195

CCC CGT GCA GCT GGA CAG CCT GGG TCA GAG CTT CTG GTG GTA ACC ATC 6794

Pro Arg Ala Ala Gly Gln Pro Gly Ser Glu Leu Leu Val Val Thr Ile

2200 2205 2210

GGG CTA GAT GGG GCC ACA CGG TTA TGG CAT CCA CTC TTG GTG TGC CAA 6842

Gly Leu Asp Gly Ala Thr Arg Leu Trp His Pro Leu Leu Val Cys Gln

2215 2220 2225

ACC CAC ACC CTC CTG GGA CAC AGC GGC CCA GTC CGT GCT GCT GCT GTT 6890
Thr His Thr Leu Leu Gly His Ser Gly Pro Val Arg Ala Ala Ala Val
2230 2235 2240 2245

TCA GAA ACC TCA GCC CTC ATG CTG ACC GCC TCT GAG ATG TCT GTA CGG 6938
Ser Glu Thr Ser Ala Leu Met Leu Thr Ala Ser Glu Met Ser Val Arg
2250 2255 2260

CTC TGG CAG GTT CCT AAG GAA GCA GAT GAC ACA TGT ATA CCA AGG AGT 6986
Leu Trp Gln Val Pro Lys Glu Ala Asp Asp Thr Cys Ile Pro Arg Ser
2265 2270 2275

TCT GCA GCC GTC ACT GCT GTG GCT TGG GCA CCA GAT GGC TCC ATG GCA 7034
Ser Ala Ala Val Thr Ala Val Ala Trp Ala Pro Asp Gly Ser Met Ala
2280 2285 2290

GTA TCT GGA AAT CAA GCT GGG GAA CTA ATC TTG TGG CAG GAA GCT AAG 7082
Val Ser Gly Asn Gln Ala Gly Glu Leu Ile Leu Trp Gln Glu Ala Lys
2295 2300 2305

GCT GTG GCC ACA GCA CAG GCT CCA GGC CAC ATA GGT GCT CTG ATC TGG 7130
Ala Val Ala Thr Ala Gln Ala Pro Gly His Ile Gly Ala Leu Ile Trp
2310 2315 2320 2325

TCC TCG GCA CAC ACC TTT TTT GTC CTC AGT GCT GAT GAG AAA ATC AGC 7178
Ser Ser Ala His Thr Phe Phe Val Leu Ser Ala Asp Glu Lys Ile Ser
2330 2335 2340

GAG TGG CAA GTG AAA CTG CGA GAG GGT TCG GCA CCC GGA AAT TTG AGT 7226
Glu Trp Gln Val Lys Leu Arg Lys Gly Ser Ala Pro Gly Asn Leu Ser
2345 2350 2355

CTT CAC CTG AAC CGA ATT CTA CAG GAG GAC TTA GGG GTG CTG ACA AGT 7274
Leu His Leu Asn Arg Ile Leu Gln Glu Asp Leu Gly Val Leu Thr Ser
2360 2365 2370

CTG GAT TGG GCT CCT GAT GGT CAC TTT CTC ATC TTG GCC AAA GCA GAT 7322

Leu Asp Trp Ala Pro Asp Gly His Phe Leu Ile Leu Ala Lys Ala Asp

2375 2380 2385

TTG AAG TTA CTT TGC ATG AAG CCA GGG GAT GCT CCA TCT GAA ATC TGG 7370

Leu Lys Leu Leu Cys Met Lys Pro Gly Asp Ala Pro Ser Glu Ile Trp

2390 2395 2400 2405

AGC AGC TAT ACA GAA AAT CCT ATG ATA TTG TCC ACC CAC AAG GAA TAT 7418

Ser Ser Tyr Thr Glu Asn Pro Met Ile Leu Ser Thr His Lys Glu Tyr

2410 2415 2420

GGC ATA TTT GTC CTG CAG CCC AAG GAT CCT GGA GTT CTT TCT TTC TTG 7466

Gly Ile Phe Val Leu Gln Pro Lys Asp Pro Gly Val Leu Ser Phe Leu

2425 2430 2435

AGG CAA AAG GAA TCA GGA AAG TTT GAA GAG AGG CTG AAC TTT GAT ATA 7514

Arg Gln Lys Glu Ser Gly Lys Phe Glu Glu Arg Leu Asn Phe Asp Ile

2440 2445 2450

AAC TTA GAG AAT CCT AGT AGG ACC CTA ATA TCG ATA ACT CAA GCC AAA 7562

Asn Leu Glu Asn Pro Ser Arg Thr Leu Ile Ser Ile Thr Gln Ala Lys

2455 2460 2465

CCT GAA TCT GAG TCC TCA TTT TTG TGT GCC AGC TCT GAT GGG ATG CTA 7610

Pro Glu Ser Glu Ser Ser Phe Leu Cys Ala Ser Ser Asp Gly Met Leu

2470 2475 2480 2485

TGG AAC CTG GCC AAA TGC AGC CCA GAA GGA GAA TGG ACC ACA GGT AAC 7658

Trp Asn Leu Ala Lys Cys Ser Pro Glu Gly Glu Trp Thr Thr Gly Asn

2490 2495 2500

ATG TGG CAG AAA AAA GCA AAC ACT CCA GAA ACC CAA ACT CCA GGG ACA 7706

Met Trp Gln Lys Lys Ala Asn Thr Pro Glu Thr Gln Thr Pro Gly Thr

2505 2510 2515

GAC CCA TCT ACC TGC AGG GAA TCT GAT GCC AGC ATG GAT AGT GAT GCC 7754

Asp Pro Ser Thr Cys Arg Glu Ser Asp Ala Ser Met Asp Ser Asp Ala

2520 2525 2530

AGC ATG GAT AGT GAG CCA ACA CCA CAT CTA AAG ACA CGG CAG CGT AGA 7802

Ser Met Asp Ser Glu Pro Thr Pro His Leu Lys Thr Arg Gln Arg Arg

2535 2540 2545

AAG ATT CAC TCG GGC TCT GTC ACA GCC CTC CAT GTG CTA CCT GAG TTG 7850

Lys Ile His Ser Gly Ser Val Thr Ala Leu His Val Leu Pro Glu Leu

2550 2555 2560 2565

CTG GTG ACA GCT TCG AAG GAC AGA GAT GTT AAG CTA TGG GAG AGA CCC 7898

Leu Val Thr Ala Ser Lys Asp Arg Asp Val Lys Leu Trp Glu Arg Pro

2570 2575 2580

AGT ATG CAG CTG CTG GGC CTG TTC CGA TGC GAA GGG TCA GTG AGC TGC 7946

Ser Met Gln Leu Leu Gly Leu Phe Arg Cys Glu Gly Ser Val Ser Cys

2585 2590 2595

CTG GAA CCT TGG CTG GGC GCT AAC TCC ACC CTG CAG CTT GCC GTG GGA 7994

Leu Glu Pro Trp Leu Gly Ala Asn Ser Thr Leu Gln Leu Ala Val Gly

2600 2605 2610

GAC GTG CAG GGC AAT GTG TAC TTT CTG AAT TGG GAA TGAAGATGTG 8040

Asp Val Gln Gly Asn Val Tyr Phe Leu Asn Trp Glu ***

2615 2620 2625

CCACTCGGGA ATAATGATAC CCCTTGTGCT AGAGATGCAA AGCCTGAAGA CACTGGTAGC 8100

TTTTAATAAT TATAAAATTA ATAATTTCTT GATAATTATA AAAATGAAGT GTCAAAAAAT 8160

CTCAAGTGTA GGCCTGCCTG TGTTCTCATG TGGATTTAGA ACAGGAGGAT ATTCTATGTG 8220

TATGTATATG TACATTCTAA TGTGTGTCTC TTCTTATTCA ACATTAATCC TTACTAGAAC 8280

CACAAGAAAG TGAATGAAAT CTTTAGTAGG TACTCTTTTG AAACTAGGTT TTAGAATTCT 8340

TGCATCACTC GCGGGCCCTA GGACCCTAGG ATGCCATTCT TGCCAGGAGG AGGAATGAGA 8400

GTGATGTTGG CCAACATTCA ATTTGAACAG AGCATGGAAG ACCTTTCAGT TCATCGGGAA 8460

AGAATGAGGG AGGGAGAATA AGTCAGTCAT GCATCAGGGC ATTTAGAAAG AGCTATGTTT 8520

CTGTCACAGA GACAGCCCTT TTCTCAGAAC TACCCAGAGG AGGCCGGGCA TGGTGGCTCA 8580

CGCTTGTAAT CCCAGCACTT TGGGAGGCCG AGGTGGGCAG ATCACGAGGT CAGGAGATCA 8640

AGACCATCCT GGCTAACATA GTGAAACCCT GTCTCTACTA AAAAATACAA AAAGTTGGCC 8700

AGGTGTGGCG GCGGGCACCT GTAGTCCCAG CTACTTGGGA GGCTGAGGCA GGAGAATGGC 8760

GTGAACCCAG GAGGCGGAGC TTGCGGTGAG CCGAGACACC ACTGCACTCC AGCCTGGGCA 8820

ACAGAGCGAG ACTCTGTCT 8839

世界知的所有権機関
国際事務局

PCT 特許協力条約に基づいて公開された国際出願

(51) 国際特許分類6 C12N 9/12, 15/54, C12Q 1/48	A1	(11) 国際公開番号	WO98/07838
		(43) 国際公開日	1998年2月26日(26.02.98)

(21) 国際出願番号 PCT/JP97/02904

(22) 国際出願日 1997年8月21日(21.08.97)

(30) 優先権データ

特願平8/219761	1996年8月21日(21.08.96)	JP
特願平9/18878	1997年1月31日(31.01.97)	JP
特願平9/31807	1997年2月17日(17.02.97)	JP

(71) 出願人（米国を除くすべての指定国について）
三菱化学株式会社
(MITSUBISHI CHEMICAL CORPORATION)[JP/JP]
〒100 東京都千代田区丸の内2丁目5番2号 Tokyo, (JP)
(72) 発明者；および
(75) 発明者／出願人（米国についてのみ）
石川冬木(ISHIKAWA, Fuyuki)[JP/JP]
〒222 神奈川県横浜市港北区篠原台町3-16-105 Kanagawa, (JP)
中村秀男(NAKAMURA, Hideo)[JP/JP]
高橋和展(TAKAHASHI, Kazuhiro)[JP/JP]
藤野泰寛(FUJINO, Yasuhiro)[JP/JP]
原田直純(HARADA, Naozumi)[JP/JP]
〒227 神奈川県横浜市青葉区鴨志田町1000番地
三菱化学株式会社 横浜総合研究所内 Kanagawa, (JP)

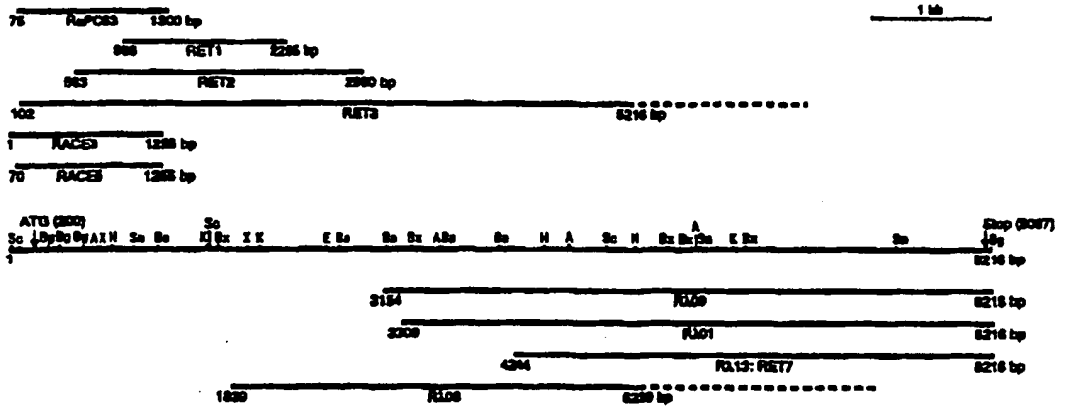
(74) 代理人
弁理士 今村正純, 外(IMAMURA, Masazumi et al.)
〒103 東京都中央区八重洲一丁目8番12号
藤和八重洲一丁目ビル7階 Tokyo, (JP)

(81) 指定国 CA, JP, US, 欧州特許 (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

添付公開書類
国際調査報告書

(54)Title: HIGHER ANIMAL TELOMERASE PROTEIN AND GENE ENCODING THE SAME

(54)発明の名称 高等動物テロメラーゼ蛋白質及びそれをコードする遺伝子



(57) Abstract

A telomerase protein originating in higher animals involving human being. This protein and a gene encoding the same are useful in, for example, the clarification of biological control mechanisms such as cell growth and aging and expected to be applicable to, in particular, the development of remedies for cancer. A method for screening substances acting on the expression of the enzyme activity of the higher animal telomerase protein involves the step of measuring the molecular weight of the telomerase protein contained in cells or tissues in contact with a test substance by, for example, the SDS polyacrylamide electrophoresis method.

（５７）要約

ヒトを含む高等動物由来のテロメラーゼ蛋白質及びそれをコードする遺伝子が提供される。テロメラーゼ蛋白質及びそれをコードする遺伝子は、例えば、細胞増殖及び細胞の老化などの生体制御機構の解明に有用であり、癌の治療薬の開発に特に有用性が期待される。また、高等動物テロメラーゼ蛋白質の酵素活性発現に作用する物質のスクリーニング方法であって、被験物質と接触させた細胞又は組織に含まれるテロメラーゼ蛋白質の分子量を例えばＳＤＳ－ポリアクリルアミド電気泳動法により測定する工程を含むスクリーニング方法が提供される。

明 細 書

高等動物テロメラーゼ蛋白質及びそれをコードする遺伝子

技術分野

本発明は高等動物細胞のテロメラーゼの蛋白質をコードする遺伝子及びその遺伝子産物に関するものである。

背景技術

動物細胞などの真核細胞染色体の線状 DNAの両末端はテロメアと呼ばれ、特殊なDNA配列とそれに結合する蛋白質からなる複雑な高次構造をとっている。テロメアDNA は、チミン(T) 及びグアニン(G) （反対鎖はアデニン(A) 及びシトシン(C)）の豊富な特徴的繰り返し配列からなり、例えば、脊椎動物細胞染色体のテロメアDNA はTTAGGG（反対鎖はCCCTAA）の６塩基の繰り返しで構成されている。この配列を利用したサザンブロッティング解析により、ヒト体細胞のテロメア繰り返し配列の平均長は７キロ～10キロベースであることが明らかにされた。

テロメア構造は染色体の安定化に重要な機能を有すると考えられている。例えば、テロメアが細胞核の辺縁に位置することが酵母を用いた形態学的研究で明かにされており、テロメアが染色体を核の特定の位置に固定する「錨」として作用し、細胞核内で各染色体間の物理的相互作用を制御している可能性が示唆されている。また、以下のように、真核細胞の線状二本鎖DNA の複製ごとの短縮化による染色体機能の不活化を防ぐ機能を有することが示唆されている。

線状二本鎖DNA の両鎖の同時複製の過程では、一方の DNA鎖（リーディング鎖）が3'末端をプライマーとして5'→3'DNA ポリメラーゼにより連続的に複製されるのに対し、他方の DNA鎖（ラギング鎖）では小さい RNAプライマーを用いた断続的なものになる。従って、新生鎖（ラギング鎖）の5'末端のRNA プライマーはDNA に置き換えられないので、細胞分裂を繰り返す毎に一方の娘細胞の5'末端が次第に短縮することになり、最後には染色体が不安定になって細胞が死に至る。しかしなが

ら、生殖細胞系列ではDNA の繰り返し複製によって染色体機能が損なわれるような染色体DNA の短縮化が生じないことが明らかにされており（Allsopp, R.C. et al., Proc. Natl. Acad. Sci. U.S.A., 89, 10114, 1992）、テロメアやそれに隣接する領域がヘアピン構造を採ったり、短縮化に対する緩衝帯として機能している可能性が示唆されている。

テロメアが染色体の短縮化を防ぐ機能を有することは、細胞の老化・不死化とテロメア繰り返し配列の平均長の変化との関係からも強く示唆されている。多細胞生物の線維芽細胞などをイン・ビトロで継代培養すると、継代を経るにつれて増殖能が低下し、最終的には増殖能を失った「老化」細胞となるが、予め細胞内にある種の癌遺伝子を導入しておくと永久増殖能を獲得した不死化細胞が得られる場合がある。これらは細胞レベル（イン・ビトロ）での老化現象及び発癌のモデルと理解されているが、分子レベルでの研究により、正常細胞では分裂回数の増加につれてテロメア繰り返し配列の平均長が短縮化し、その平均長は継代可能回数と相関すること、並びに、不死化細胞ではテロメア繰り返し配列の平均長が短いが、継代中にその平均長が変化しないことが明らかにされた。

テロメア繰り返し配列平均長の制御機構の一つとして、テロメア繰り返し配列を伸長させる RNA依存性 DNAポリメラーゼ（テロメラーゼ）が注目されている。この酵素は、原生動物テトラヒメナの大核抽出液中から、テトラヒメナのテロメア繰り返し配列由来の合成オリゴヌクレオチド(TTGGGG)の3'端に同じ６塩基の繰り返し配列を付加する酵素として見い出されたものであり、活性に必要なサブユニットとしてテロメアDNA 配列の5'-TTAGGG-3'に相補的な鋳型RNA を含み、鋳型RNAを基にしてテロメアDNA の一本鎖を延長する一種の逆転写酵素である。テトラヒメナ・テロメラーゼ由来のテロメラーゼが精製され、そのcDNAがクローニングされた（Collins, K, et al., Cell, 81, 677, 1995)。このテロメラーゼは鋳型RNA と結合する80 kD のサブユニット及びプライマーとなるDNA 末端に結合する95 kD のサブユニットからなり、RNA ウイルスの RNAポリメラーゼに比較的類似の一次構造を有することが明かにされた。

テロメラーゼの生物学的意義は、テトラヒメナや酵母などの下等真核生物で明

らかにされた。すなわち、テトラヒメナ・テロメラーゼRNA 遺伝子のテロメア繰り返し配列の鋳型部分に点突然変異を導入した遺伝子で形質転換された個体では、導入されたある種の点突然変異に対応する変異テロメア繰り返し配列が生合成されると同時に増殖不可能になる。また、パン酵母・テロメラーゼ RNA遺伝子である TLC1 が破壊されると、継代を重ねるにつれてその酵母のテロメア繰り返し配列平均長が短くなり、最終的には増殖不可能となる。従って、単細胞真核生物ではテロメラーゼが細胞増殖に必須の酵素であると理解されている。

イン・ビトロでのヒト細胞の不死化過程において、テロメラーゼ活性が癌遺伝子導入後の継代初期には認められず、無限増殖能を獲得した細胞集団において検出されることが明らかにされた。また、実際のヒト癌細胞のほとんどにテロメラーゼ活性が検出される一方で、多くの正常細胞ではテロメラーゼ活性は検出されないと言われている。これらの知見から、癌細胞は、その成立過程においてテロメラーゼ活性の発現によりテロメアDNA の短縮化を免れ、永久増殖能を獲得するのではないかとの推測が可能である。従って、テロメラーゼ阻害剤が選択性の高い抗癌剤として有用であり、テロメラーゼ活性の測定により癌の早期診断が可能になると予測される。

テロメラーゼＲＮＡサブユニットの発現の程度は必ずしもテロメラーゼ活性に相関しないという報告がある（Ａｖｉｌｏｎら、Ｃａｎｃｅｒ　Ｒｅｓ．，５６、６４５、１９９６）。しかしながら、現在のところ、ヒトを含めた高等動物においてはテロメラーゼ自体が未だ分離・精製されておらず、その物質的実態は不明のままであり、しかも、実際にテロメラーゼ活性を検出するためにはＰＣＲを用いた煩雑な検出法を用いる必要があるので、テロメラーゼについての酵素学的研究はほとんどなされていないのが現状である。さらに、病理切片などを用いてテロメラーゼ活性の発現を個々の細胞レベルで判定することもできないため、テロメラーゼと癌の悪性度との正確な関係を解析することは困難である。

従って、テロメラーゼ蛋白質を単離・同定することによって、高等動物テロメラーゼの物質的特徴を解明するとともに、酵素学的見地からテロメラーゼの阻害剤の研究を行い、テロメラーゼと癌の悪性度との関係を解明することが強く望ま

れている。

発明の開示

そこで本発明者らは、高等動物テロメラーゼ蛋白質を単離・同定するべく鋭意検討を重ね、高等動物テロメラーゼ蛋白質をコードする遺伝子のクローニングに初めて成功し、さらにその遺伝子から遺伝子産物である高等動物テロメラーゼ蛋白質を発現させることに成功した。また、この遺伝子産物を特異的に認識する抗体を作製し、これを用いてテロメラーゼ活性とこの遺伝子産物との密接な関係を証明することにも成功した。本発明はこれらの知見を基にして完成されたものである。なお、最近、ヒト・テロメラーゼ蛋白質の全長のアミノ酸配列が報告されたが (Science, 275, pp.973-977, February 14, 1997)、c-DNA の塩基配列及びアミノ酸配列は本発明者らが解明したものと多くの部分で相違している。

本発明は、配列表の配列番号１に記載のアミノ酸配列で特定されるポリペプチドを提供するものであり、該ポリペプチドはラット由来テロメラーゼ蛋白質であることを特徴としている。また、本発明により、配列表の配列番号１に記載のアミノ酸配列に１又は２以上のアミノ酸残基による置換、挿入、及び／又は欠失が存在しており、実質的にヒトを含む高等動物テロメラーゼ蛋白質として機能することを特徴とするポリペプチドが提供され、その好ましい態様により、ヒトの生体内でテロメラーゼ蛋白質として機能することができる上記ポリペプチドが提供される。

また、本発明の別の態様により、配列表の配列番号２に記載のアミノ酸配列で特定されるポリペプチドが提供されるが、このポリペプチドはヒト由来テロメラーゼ蛋白質の部分ポリペプチドであることを特徴としている。さらに本発明により、配列表の配列番号２に記載のアミノ酸配列に１又は２以上のアミノ酸残基による置換、挿入、及び／又は欠失が存在しており、実質的にヒトを含む高等動物テロメラーゼ蛋白質の部分ポリペプチドとして機能することを特徴とするポリペプチドが提供される。

さらに本発明の別の態様により、配列表の配列番号１３に記載のアミノ酸配列

で特定されるポリペプチドが提供されるが、該ポリペプチドはヒト由来テロメラーゼ蛋白質であることを特徴としている。また、本発明により、配列表の配列番号１３に記載のアミノ酸配列に１又は２以上のアミノ酸残基による置換、挿入、及び／又は欠失が存在しており、実質的にヒトを含む高等動物テロメラーゼ蛋白質として機能することを特徴とするポリペプチドが提供され、その好ましい態様により、ヒトの生体内でテロメラーゼ蛋白質として機能することができる上記ポリペプチドが提供される。

さらに本発明の別の態様によれば、上記の各ポリペプチドをコードするヌクレオチド配列が提供される。このヌクレオチド配列としては、ＤＮＡ配列又はＲＮＡ配列を挙げることができ、例えば、その好ましい態様として、配列表の配列番号１に記載のＤＮＡ配列の核酸番号１９９から核酸番号８０８５（終始コドンを含まず）で特定されるＤＮＡ、又は配列表の配列番号２に記載のＤＮＡ配列の核酸番号１から核酸番号４８７で特定されるＤＮＡ、又は配列表の配列番号１３に記載のＤＮＡ配列の核酸番号１５６から核酸番号８０３０（終始コドンを含まず）で特定されるＤＮＡが提供される。以上に加えて、上記ＤＮＡ配列を含む組み換えベクター、該組み換えベクターが導入された形質転換体、及び、該形質転換体を培養した培養物から上記ＤＮＡ配列の遺伝子産物であるポリペプチドを分離・採取する工程を含む、上記ポリペプチドの製造方法も提供される。

本発明のさらに別の態様として、上記の各ポリペプチドを特異的に認識することができる抗体、上記の各ヌクレオチド配列の一部又は全部に相補的に結合可能なヌクレオチド配列を含む核酸プローブが提供されるが、これらの抗体又は核酸プローブは癌細胞検出用試薬として有用であり、上記抗体又は核酸プローブを含む癌診断用の医薬組成物が本発明の一態様として提供される。

これらの発明に加えて、本発明の別の態様により、ＳＤＳ（ドデシル硫酸ナトリウム）－ポリアクリルアミド電気泳動法（ＰＡＧＥ）による分子量が、不活性型では約２４０ｋＤａであり、活性型では約２３０ｋＤａであることを特徴とする上記ポリペプチドと、ＳＤＳ－ポリアクリルアミド電気泳動法による分子量が約２３０ｋＤａであることを特徴とする活性型のポリペプチドが提供される。ま

た、高等動物テロメラーゼ蛋白質の酵素活性の発現に作用する物質のスクリーニング方法であって、被験物質と接触させた細胞又は組織に含まれる高等動物テロメラーゼ蛋白質のサブユニットであるポリペプチドの分子量を測定する工程を含む方法も提供される。

上記方法の発明の好ましい態様によれば、被験物質との接触工程を被験物質の存在下における培養工程又は動物への被験物質の投与工程により行う上記方法；分子量の測定をＳＤＳ－ポリアクリルアミド電気泳動法で行う上記方法；約２４０ｋＤａの不活性型ポリペプチド及び約２３０ｋＤａの活性型のポリペプチドの存在比を測定する工程を含む上記方法；被験物質の非存在下における２４０ｋＤａのポリペプチドの存在比と比較して、該ポリペプチドの存在比が被験物質の存在下において実質的に増加している場合には、該被験物質が高等動物テロメラーゼ蛋白質の酵素活性の発現を阻害する物質であると判定する工程を含む上記方法；並びに、被験物質の非存在下における２３０ｋＤａのポリペプチドの存在比と比較して、該ポリペプチドの存在比が被験物質の存在下において実質的に増加している場合には、該被験物質が高等動物テロメラーゼ蛋白質の酵素活性の発現を活性化する物質であると判定する工程を含む上記方法が提供される。

図面の簡単な説明

第１図は、ラット・テロメラーゼ蛋白質遺伝子のｃＤＮＡクローンの制限酵素切断地図を示した図である。

第２図は、ＰＣＲによって増幅されたヒト・テロメラーゼ蛋白質遺伝子のｃＤＮＡ断片のＤＮＡ配列と、予想されるアミノ酸配列について、それぞれラットのもの又はテトラヒメナｐ８０との相同性を比較した結果を示した図である。図中、Ｒはラット遺伝子、Ｈはヒト遺伝子、ｐ８０はテトラヒメナｐ８０遺伝子を示す。

第３図は、組み換えラット・テロメラーゼ蛋白質断片に対する特異抗体をコートしたビーズを用いて、ヒト癌細胞（ＰＡ－１）またはラット癌細胞（ＡＨ６６Ｆ）抽出液由来のテロメラーゼ活性が免疫沈降させた結果を示した図である。ＰＣＲとＥＬＩＳＡを組み合わせた方法を用いて検討した結果を示してあり、縦軸はテ

ロメラーゼ活性を表し、「ビーズのみ」は抗体をコートしていない陰性対照、「ＰＩ－１」は免疫前血清由来ＩｇＧをコートした陰性対照を示す。「１－４１ｄ」と「Ｒ１－１１６ｄ」は過免疫血清由来特異ＩｇＧをコートしたサンプルの結果を示す。

第４図は、ヒト・テロメラーゼ蛋白質遺伝子のｃＤＮＡクローンの制限酵素切断地図を示した図である。

発明を実施するための最良の形態

本発明のポリペプチドの第一の態様は、配列表の配列番号１に記載のアミノ酸配列で特定され、マウス由来のテロメラーゼ蛋白質を構成するポリペプチドに相当するものである。本発明により提供される上記ポリペプチドは、配列番号１に記載された特定のポリペプチドに限定されることはなく、配列表の配列番号１に示されたアミノ酸配列に１又は２以上のアミノ酸残基による置換、挿入、及び／又は欠失が存在しており、ヒトを含む高等動物のテロメラーゼ蛋白質として実質的に機能することができるポリペプチドも本発明の範囲に包含される。また、このようなポリペプチドをサブユニットとして含む高等動物テロメラーゼ蛋白質も本発明の範囲に包含される。

本発明のポリペプチドの第二の態様は、配列表の配列番号２に記載のアミノ酸配列で特定され、ヒト由来のテロメラーゼ蛋白質を構成するポリペプチドの部分ポリペプチドに相当するものである。本発明により提供される上記ポリペプチドは、配列番号２に記載された特定のポリペプチドに限定されることはなく、配列表の配列番号２に示されたアミノ酸配列に１又は２以上のアミノ酸残基による置換、挿入、及び／又は欠失が存在しており、実質的に高等動物、好ましくはヒトのテロメラーゼ蛋白質の部分ポリペプチドとして機能することができるポリペプチドも本発明の範囲に包含される。

本発明のポリペプチドの第三の態様は、配列表の配列番号１３に記載のアミノ酸配列で特定され、ヒト由来のテロメラーゼ蛋白質を構成するポリペプチドに相当するものである。本発明により提供される上記ポリペプチドは、配列番号１３に

記載された特定のポリペプチドに限定されることはなく、配列表の配列番号１３に示されたアミノ酸配列に１又は２以上のアミノ酸残基による置換、挿入、及び／又は欠失が存在しており、ヒトを含む高等動物のテロメラーゼ蛋白質として実質的に機能することができるポリペプチドも本発明の範囲に包含される。また、このようなポリペプチドをサブユニットとして含む高等動物テロメラーゼ蛋白質も本発明の範囲に包含される。

本発明のポリペプチドには、上記の各ポリペプチドを部分配列として含むポリペプチドも包含される。例えば、上記の各ポリペプチドに対してその発現効率を向上させる性質を有する適宜のアミノ酸配列を結合させたポリペプチド、上記の各ポリペプチドに対してシグナル配列を結合させたポリペプチド、上記ポリペプチドの発現を確認するために読み枠が変わらないように他の蛋白質と上記ポリペプチドとを結合させた、いわゆるタグ配列との融合蛋白質なども本発明の範囲に包含される。

上記のポリペプチドのうちのいずれかをコードするヌクレオチド配列は、いずれも本発明のヌクレオチド配列に包含される。本発明のテロメラーゼ蛋白質をコードする遺伝子（本明細書において「テロメラーゼ蛋白質遺伝子」という場合があり、テロメラーゼ蛋白質を構成するポリペプチドの全長又はその一部をコードするヌクレオチド配列を意味するものとして用いる）としては、上記の第一の態様、第二の態様、及び第三の態様に包含されるポリペプチドをコードするヌクレオチド配列、好ましくはＤＮＡ配列を挙げることができる。

本明細書において「高等動物」という用語は、ヒトを含む哺乳類動物を包含する概念として用いる。このような高等動物、好ましくは哺乳類動物に由来するテロメラーゼ蛋白質を構成するポリペプチドは、それぞれ高い相同性を有していることが期待される。従って、本明細書に詳細に開示されたマウス由来のテロメラーゼ蛋白質遺伝子についてのクローニング方法及びその遺伝子の情報を基にすれば、当業者は高等動物由来のテロメラーゼ蛋白質を構成するポリペプチドをコードする遺伝子を容易に入手できるとともに、その遺伝子産物を取得することが可能であることはいうまでもない。

本発明のテロメラーゼ蛋白質遺伝子は、例えば次のような方法によって得られる。本発明のテロメラーゼ蛋白質遺伝子を含有するＤＮＡライブラリーとしては、不死化した高等動物細胞株、好ましくはヒト、サル、ウマ、ウシ、ヒツジ、イヌ、ネコ、ウサギ、ラット、マウスなどの細胞株から調製したＲＮＡを用いて公知の常法により作成したプラスミドｃＤＮＡライブラリー若しくはファージｃＤＮＡライブラリー、又はファージゲノミックライブラリーなどが利用できる。

例えば、ファージｃＤＮＡライブラリーを用いる場合には、まず、癌などの組織、あるいは不死化した高等動物細胞株を液体窒素中で粉砕し、グアニジンイソチオシアネート水溶液等中でホモジナイズした後、Ｃｈｉｒｇｗｉｎらの方法［Ｂｉｏｃｈｅｍｉｓｔｒｙ　１８、５２９４－５２９９（１９７９）］に従って塩化セシウム平衡密度勾配遠心法によって全ＲＮＡを沈殿として分離する。ＲＮＡの分離には市販のＲＮＡｚｏｌ（ＴｅｌＴｅｓｔ社）などの抽出試薬を使用することもできる。ＲＮＡの分離後、フェノール抽出、エタノール沈殿により全ＲＮＡを精製し、オリゴ（ｄＴ）セルロースカラムクロマトグラフィーに付して精製することにより、目的のテロメラーゼ蛋白質のｍＲＮＡを含むポリ（Ａ）含有ｍＲＮＡ（ｐｏｌｙA^+ ｍＲＮＡ）群を調製することができる。

次に、上記で調製したｍＲＮＡ群に対して、例えば、デオキシチミジンが１２個から１８個連結したいわゆるＯｌｉｇｏ（ｄＴ）配列自体、又はネーチャー［Ｎａｔｕｒｅ　３２９、８３６－８３８（１９８７）］に記載されているようなＯｌｉｇｏ（ｄＴ）配列を含有するような合成ＤＮＡにより構成されるプライマーＤＮＡをハイブリダイズさせ、逆転写酵素により１本鎖ｃＤＮＡを合成する。市販のｃＤＮＡの合成キットにもこれに類する配列が利用されているので、そのような配列を用いてもよい。その後、市販プライマーに対するＰＣＲ反応用の合成ＤＮＡ（通常はキットに添付されているもの自体）を用いてＰＣＲ反応を行えば良い。また、前記文献［Ｎａｔｕｒｅ　３２９、８３６.－８３８（１９８７）］に記載されているようなプライマーＤＮＡを用いる場合には、その配列に相補的な配列を設計し、ＰＣＲ反応用のプライマーとしてあらかじめ用意しておくことが好ましい。その後、大腸菌のＤＮＡポリメラーゼＩ、大腸菌のＤＮＡリガーゼ、

ＲＮａｓｅＨを用いて、常法に従って２本鎖ｃＤＮＡを合成する。次いで、Ｔ４ＤＮＡポリメラーゼによりｃＤＮＡの末端を平滑化した後、いわゆるＥｃｏＲＩアダプター等の、制限酵素により切断された形をなすＤＮＡの小断片をＴ４ＤＮＡリガーゼによりｃＤＮＡ鎖の両末端に付加する。

この際、例えばＥｃｏＲＩメチレース等のＤＮＡメチレースでｃＤＮＡ中の制限酵素切断点をメチル化し（例えば、ＥｃｏＲＩメチレースの場合はＥｃｏＲＩ切断点のメチル化を行い）、制限酵素ＥｃｏＲＩの切断からｃＤＮＡを保護しておき、次に、ｃＤＮＡの末端に、いわゆるＥｃｏＲＩリンカー等をＴ４ＤＮＡリガーゼにより付加した後、制限酵素ＥｃｏＲＩでリンカーＤＮＡ部分のみを切断しても同様な結果が得られる。ベクターのクローニングサイトとして、例えばＢａｍＨＩなどの他の制限酵素の切断点を選択する場合には、前述の一連の末端処理の操作を、例えばＢａｍＨＩアダプターの結合もしくはＢａｍＨＩメチレース、ＢａｍＨＩリンカー、ＢａｍＨＩ等の組み合わせで処理にすることによっても同様な結果を得ることができる。

上記の様に末端処理されたｃＤＮＡ鎖を市販のλファージベクター、例えばλＺＡＰ（ＰｒｏｍｅｇａＢｉｏｔｅｃｈ社）等のλファージベクターまたはｐＧＥＭ２（ＰｒｏｍｅｇａＢｉｏｔｅｃｈ社）等のプラスミドベクターのＥｃｏＲＩ切断部位に常法に従って挿入することにより、組換えλファージＤＮＡ群または組換えプラスミドＤＮＡ群を製造することができる。あるいは、ＰＣＲ反応を用いて断片を取得する場合には、ＰＣＲ反応により増幅されたＤＮＡの断片の末端に特異的に〔Ａ〕が付加されるために、それに相補的な〔Ｔ〕を付加したベクター、例えばｐＣＲＩＩ（Ｉｎｖｉｔｒｏｇｅｎ社）やｐＴ７（Ｎｏｖａｇｅｎ社）などのベクターを用いて製造することができる。

このようにして得られた組換えλファージＤＮＡ群を材料として、市販のイン・ビトロ・パッケージング・キット、例えばギガパック・ゴールド（プロメガ・バイオテック社）などを用いていわゆるイン・ビトロ・パッケージングを行い、組換えλファージＤＮＡを有するλファージ粒子を製造することができる。パッケージングは、一般には、市販のキットの添付説明書の条件に従って行えばよい。得

られたλファージ粒子を常法、例えばT. Maniatisらの方法（「Molecular Cloning」、Cold Spring Harbor Laboratories 1982年）に従い、例えば大腸菌などの宿主に形質導入し、得られた形質転換体を増殖させることによってファージcDNAライブラリーを作ることができる。また、組換えプラスミドDNA群では、常法に従い、例えば大腸菌などの宿主に形質転換し、得られた形質転換体を増殖させことによって、プラスミドcDNAライブラリーを得ることができる。

次に、これらファージあるいは大腸菌などの形質転換体を増殖させ、例えばジーンスクリーンプラス（Dupont社）などのナイロン膜あるいはニトロセルロース膜上に移し取り、アルカリ存在下で蛋白を除くことにより調製したλファージDNAあるいはプラスミドDNAに対して、後述の方法で増幅された高等動物テロメラーゼ蛋白質遺伝子の部分断片から作製した［^{32}P］標識プローブをハイブリダイズさせ、プラークハイブリダイゼーション法によって選択し、目的とする高等動物テロメラーゼ蛋白質遺伝子をコードするcDNAクローンの全部または一部を得ることができる。

ファージcDNAライブラリーまたはプラスミドcDNAライブラリーから目的とする高等動物テロメラーゼ蛋白質遺伝子をコードするcDNAクローンを選択する為に用いるプローブは、常法に従い、例えば市販のキット等を用いて調製することができる。例えば、既知のテロメラーゼ蛋白質（Collinsら、Cell、81、677－686、1995）をコードする遺伝子に由来するDNA配列や、そのアミノ酸配列と相同性を有するアミノ酸配列をコードし得る別の生物の遺伝子のDNA配列をNational Center for Biotechnology Information（NCBI）などの遺伝子バンク中でTBLASTNなどのプログラムを用いて検索し、ある程度相同性を有するアミノ酸配列について、それをコードし得るDNA配列を参考にしてオリゴヌクレオチドを合成してプローブとして用いることができる。また、同様な遺伝子のDNA配列を基にPCRプライマーを設計し、いわゆるdegenerative PCR法によって、より長いDNAを取得してプロー

プとして用いてもよい。この場合、ＰＣＲ法に用いる鋳型には、目的のプローブＤＮＡを含む細胞由来のファージｃＤＮＡライブラリー、プラスミドｃＤＮＡライブラリー、または抽出したＲＮＡから常法に従って合成したｃＤＮＡなどを用いることができる。

また、上記のように遺伝子ライブラリーをハイブリダイゼーション法でスクリーニングせずに、プローブＤＮＡを設計したようにＰＣＲプライマーを設計し、いわゆるＰＣＲ法で高等動物のテロメラーゼ蛋白質遺伝子の一部を取得することもできる。その場合、ＰＣＲ法に用いる鋳型としては、前述のファージｃＤＮＡライブラリー、プラスミドｃＤＮＡライブラリーの他、不死化細胞より抽出したＲＮＡから常法に従って合成したｃＤＮＡを直接用いることができる。ＰＣＲ反応後、反応液をアガロースやポリアクリルアミドゲル電気泳動で解析し、二種類のプライマーにより増幅されるＤＮＡ断片の中から、予想される大きさの断片を回収、精製し、例えばｐＣＲ－ＩＩの様なＰＣＲ断片を直接組み込むことができる市販のベクターに結合し、得られた組み換えベクターで大腸菌などの宿主を形質転換して塩基配列の解析に用いることができる。さらに、得られた高等動物テロメラーゼ蛋白質遺伝子の部分配列を基にして新たにＰＣＲプライマーを設計、合成し、高等動物テロメラーゼ蛋白質の配列を基に設計したＰＣＲプライマー、あるいはｃＤＮＡを合成する際に用いるプライマーに対して相補的な配列のプライマー、またはｃＤＮＡの両端に付加したアンカー配列に対応するＰＣＲ用プライマー、ｃＤＮＡが組み込まれたベクターに対するプライマーと、新たに合成した上記プライマーとの間でＤＮＡの増幅を繰り返し行うことによって高等動物テロメラーゼ蛋白質の全長をコードする遺伝子を取得することもできる。

ＰＣＲ反応の終了後、ＤＮＡの断片をアガロース又はポリアクリルアミドゲル電気泳動に付して常法に従って解析、回収、及び精製を行うことができる。得られた精製ＤＮＡ断片を、例えばｐＣＲ－ＩＩの様なＰＣＲ断片を直接組み込むことができるベクターに挿入し、得られた組み換えベクターで大腸菌を形質転換して常法に従ってＤＮＡを調製し、Ｓａｎｇｅｒらのジデオキシ法［Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、７４、５４６３、１９７７年］によって

目的ＤＮＡ断片の塩基配列を決定することができる。配列の決定はＡＢＩ３７３Ａ（アプライド・バイオ・システムズ社）の様な自動シークエンサーによって行うこともできる。

またファージライブラリーやプラスミドライブラリーから得られたクローンの場合、一般的には、自動シークエンサーを用いて塩基配列を決定できる配列長には限界があるため、ベクターに挿入されたｃＤＮＡの全領域を一度に解析することが困難な場合がある。このような場合には、断片を適当な制限酵素で切断した後、断片をゲル電気泳動で分離、回収し、さらに回収した断片を適宜のベクターに挿入し直すことにより解析を容易にすることができる。このような操作（サブクローニング）の他、自動シークエンサーが決定した塩基配列の中から適当な配列を選び、新たなプライマーを設計して、そこから先を継続して解析することもできる。このようにして決定されるＤＮＡ断片の配列を互いに重なるようにつなぎ合わせることにより、例えば、配列表の配列番号１または１３に記載したような高等動物テロメラーゼ蛋白質を構成する全長ポリペプチドをコードするヌクレオチド配列、又は配列表の配列番号２に記載したような高等動物テロメラーゼ蛋白質を構成する部分ポリペプチド配列をコードするヌクレオチド配列を決定することができる。

本発明のヌクレオチドにはＤＮＡ及びＲＮＡが包含されるが、配列表の配列番号１、１３、及び２には、それぞれ、ラット及びヒト由来テロメラーゼ蛋白質を構成する全長ポリペプチドをコードするＤＮＡ配列、並びにヒト由来テロメラーゼ蛋白質を構成する部分ポリペプチド配列をコードするＤＮＡ配列を好ましい態様として記載した。本発明のヌクレオチドには、上記の配列番号１、１３、及び２により特定されるＤＮＡ配列のほか、それらがコードするポリペプチドのアミノ酸配列に対して１又は２以上のアミノ酸残基による置換、挿入、及び／又は欠失が導入されており、実質的に高等動物テロメラーゼ蛋白質の全長又は部分ポリペプチドとして機能するポリペプチドをコードするヌクレオチドが包含される。このようなアミノ酸残基の置換、挿入、及び／又は欠失等によるアミノ酸配列の改変は、例えば、Ｎｕｃｌｅｉｃ Ａｃｉｄ Ｒｅｓ．，Ｖｏｌ．１０．６４８７

－６５００（１９８２）、Ｍｅｔｈｏｄｓ ｉｎ Ｅｎｚｙｍｏｌ．，Ｖｏｌ．２１７，２１８－２２７（１９９３），同Ｖｏｌ．２１７，２７０－２７８（１９９３）等に記載の部位特異的変異技術により行うことができるが、これらの方法に限定されることはなく、当業者に利用可能なものであればいかなる方法を用いてもよい。

以上のようにして得られた高等動物テロメラーゼ蛋白質遺伝子ＤＮＡの少なくとも一部分をハイブリダイゼーション・プローブまたはＰＣＲプライマーとして用いることにより、他の種の高等動物テロメラーゼ蛋白質遺伝子を同様な方法で単離することができる。例えば、テトラヒメナ・テロメラーゼ蛋白質（ｐ８０）とラット・テロメラーゼ蛋白質のアミノ酸配列の相同性の最も高い部分に由来するＰＣＲプライマーを用いて、対応する部分のヒト・テロメラーゼ蛋白質のアミノ酸配列を明らかすることも可能であり、さらにはその全長ｃＤＮＡを得ることもできる。

上記のようにして得られる高等動物テロメラーゼ蛋白質遺伝子ＤＮＡ又はそのＤＮＡ断片は、その両端あるいはどちらか一端を改変し、またはそれ自体で、公知の発現ベクターにそれ自体公知の方法でプロモーターの下流に挿入することができ、このようにして製造される遺伝子発現用の組み換えベクターを、大腸菌、酵母、動物細胞宿主等、公知の細胞中にそれ自体公知の方法により導入して形質転換体を製造することができる。

本発明の高等動物テロメラーゼ蛋白質の産生方法につき詳細に説明すると、発現用ベクターとしては、上記のようにして得られた高等動物テロメラーゼ蛋白質をコードするＤＮＡを転写できる位置にプロモーターを含有しているものが使用される。

高等動物テロメラーゼ蛋白質の工業的生産のためには、安定した宿主－ベクター系を構築すること、さらに生物学的に活性を有する高等動物テロメラーゼ蛋白質を発現しうる系を用いる必要がある。高等動物テロメラーゼ蛋白質は比較的大きな蛋白質であり、そのリフォールディングが生理活性の獲得に重要である。一般的には、リフォールディングを考慮した場合、宿主として動物細胞を用いること

が有利である。高等動物テロメラーゼは、数種の蛋白質及びＲＮＡサブユニットからなる複合体として存在する可能性があり、生理活性のある高等動物テロメラーゼとして組み換え体から精製する場合には、導入する高等動物テロメラーゼ蛋白質の由来する生物種と宿主細胞の由来する生物主の一致することが好ましい。もっとも、高等動物テロメラーゼ蛋白質を大腸菌で生産させた後、活性を有する複合体としてｉｎ　ｖｉｔｒｏで他の構成成分と再構成することが可能であることはいうまでもない。

動物細胞としては、例えばＣＨＯ細胞（生物種：ハムスター）、ＣＯＳ細胞（生物種：サル）、ＮＩＨ３Ｔ３細胞（生物種：マウス）、Ｒａｔ－１（生物種：ラット）細胞、ＶＡ－１３（生物種：ヒト）細胞等が挙げられる。これらの細胞を宿主とした発現用プラスミドは、プロモーターとしてはＳＶ４０プロモーター由来またはウイルス遺伝子由来のプロモーターが好ましい。この下流に高等動物テロメラーゼ蛋白質遺伝子を５’側から挿入する。また高等動物テロメラーゼ蛋白質の生産量を上げるために、高等動物テロメラーゼ蛋白質遺伝子を５’側から２～３個つなげたものを挿入してもよいし、各高等動物テロメラーゼ蛋白質遺伝子の５’側にＳＶ４０などのプロモーターを挿入したものを２～３個つなげてもよい。この高等動物テロメラーゼ蛋白質遺伝子の下流にポリアデニル化部位を含むことが好ましく、例えばＳＶ４０ＤＮＡ、β－グロビン遺伝子またはメタロチオネイン遺伝子由来のものを用いることができる。

このような発現ベクターは、例えばＣＨＯ細胞などの動物細胞に形質転換した際の選択マーカーを有していてもよい。選択マーカーを用いる場合には、例えば、メトトレキセート耐性を与えるＤＨＦＲ遺伝子、ネオマイシン誘導体Ｇ－４１８耐性遺伝子などを用いることができる。各耐性遺伝子の５’側に例えばＳＶ４０由来のプロモーターが挿入されており、各耐性遺伝子の３’側にポリアデニル化部位が含まれていることが好ましい。高等動物テロメラーゼ蛋白質の発現ベクターに対してこれらの耐性遺伝子を挿入する場合、高等動物テロメラーゼ蛋白質遺伝子のポリアデニル化部位下流に挿入すればよい。また、発現ベクターは形質転換体の選択マーカーを有していなくてもよい。この場合には、高等動物テロメラー